UNIVERSITÉ DE MONTRÉAL

CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, SUCCURSALE A, MONTREAL, QUEBEC, H3C 3J7

ON THE ESTIMATION OF MODELS WITH GENERALIZED SAR(1) PROCESSES ON THE RESIDUALS OF A REGRESSION

by

Denis Bolduc

I greatly benefited from discussions with Marcel G. Dagenais, Christian Gourieroux, Marc Gaudry and Moshe Ben-Akiva.

The research for this paper was supported by the F.C.A.R. program of Quebec and S.S.h.R.C. of Canada.

Cette étude a été publiée grâce à une subvention du fonds F.C.A.R. pour l'aide et le soutien à la recherche.

Université de Montréal
Centre de recherche sur les transports - Publication #508
January 1987

Dépôt légal - 1er trimestre 1987

ABSTRACT

We propose a generalized approach for dealing with spatially autocorrelated residuals. It is a generalization of a SAR(1) model [spatial autoregressive of order 1] where the weights w_{ij} are being estimated. In previous applications of SAR(1) formulations, these weights have always been fixed before estimation. Here the w_{ij}'s are specified as functional forms that can be estimated. A maximum Likelihood estimation procedure is used. The conditions under which the estimations are consistent and asymptotically normally distributed are described. In the last part of the paper, an application of the method is presented. Since the concept of spatial autocorrelation is for many, quite unkown the most important results previously obtained in the literature are summarized in the first part of the paper.

Keywords : Spatial Autocorrelation, Maximum Likelihood Estimation, Consistency of Estimation.

RESUME

Afin de contrecarrer les effets de l'autocorrelation spatiale dans les residus de regression, une approche SAR(1) generalisee, ou les poids w_{ij} sont estimes, est proposee. Dans la pratique courante d'estimation en presence d'autocorrelation spatiale, les poids w_{ij} sont fixes avant estimation. Ici, ce sont des formes fonctionelles estimables conjointement avec les parametres de la regression. Une methode du maximum de vraisemblance est utilisee. Les conditions sous lesquelles les estimations sont convergentes et asymptotiquement normalement distribuees sont decrites. La section A contient une revue breve de la litterature, tandis que la section C est consacree a l'etude des resultats d'une application de l'approche.

Mots cles : autocorrelation spatiale, estimation du maximum de vraisemblance, convergence des estimateurs.

INTRODUCTION

When applying the concept of autocorrelation to the residuals of a regression that uses cross-sectional data, one must allow for the possibility that the residuals ϵ_i of area i may be correlated to the error ϵ_j $(j \neq i)$ of other areas of the study. More precisely, because common unobserved factors may be shared by neighboring areas, the error term ϵ_i of area i $(i=1,\ldots,N)$ can often be expected to be correlated with the error of contiguous areas. In the literature, such residuals are said to be spatially autocorrelated.

In order to treat the problem of spatial correlation in the residuals we consider here a generalized specification of the spatial autoregressive of order one "SAR(1)" formulation. According to this formulation, which belongs to a wider class of spatial models called STARIMA [Space Time Autoregressive and Integrated Moving Average] initially proposed by Cliff and Ord (1975), the error ϵ_i is expressed as a weighted sum over the other error terms. Analytically:

$$\epsilon_i = \rho \sum_{j=1}^{N} w_{ij}\, \epsilon_j + u_i \qquad i=1,\ldots,N \qquad \text{with } w_{ii}=0,$$

where ρ is a spatial autocorrelation coefficient that captures the strength and the sign of the correlation between the residuals. With positive spatial autocorrelation $(\rho > 0)$, similar values of the residuals tend to be bunched together. In the case of negative correlation $(\rho < 0)$, the similar values of the residuals tend to be scattered over the region. To leave to ρ the ability to depict the sign of the autocorrelation, the weights w_{ij}'s are postulated to be non-negative. The weights w_{ij} are some functional forms specified to take into consideration the contiguity structure and the economico-geographical character of the region. For example the fact that nearby areas are more likely to be correlated than areas that are far apart may be modelled by specifying $w_{ij} = d_{ij}^{-\alpha}$ with $\alpha > 0$. For a fixed value of α, the greater the distance between two areas is, the less they are related. More generally these functions can be written as:

$$w_{ij} = g(V_{ij}^{1}, V_{ij}^{2}, \ldots, V_{ij}^{P}, \theta_1, \theta_2, \ldots, \theta_H),$$

where there is H parameters θ_h attached to P variables that can explain the spatial correlation in the residuals. A commonly employed form called boolean sets $w_{ij} = 1$ if area i is physically contiguous to area j and 0 otherwise.

In previous empirical applications of this approach, the parameters of the w_{ij}'s have always been fixed before estimation. As a generalization of the SAR(1) residual formulation, we describe in section B of this paper a methodology that jointly estimates ρ, the parameters of the w_{ij} and those of the regression. The method of estimation uses the maximum likelihood principle. To better understand the complications that can occur in this joint estimation problem, we first investigate in section A the problems related to the estimation of ρ. Then the condition under which the maximum likelihood estimators for this model are consistent and asymptotically normally distributed are described. Some applications are produced in section C.

The main purpose of this paper is to provide econometricians with a self-contained presentation of some previous and new results related to the estimation of regressions with spatially autocorrelated residuals. Since many econometricians are either unaware or unfamiliar with this topic, some examples and the most important results previously considered in the literature are included. It is worth mentioning that over the last 25 years, this problem has almost exclusively been treated by geographers and statistician geographers. Recent contributions closely related to the topic considered here are Hepple (1979), Haining (1980) and Anselin (1982). Also noteworthy are papers dedicated to developping statistical tools for testing the presence of spatial autocorrelation. These are: Cliff and Ord (1969), Brandsma and Ketellaper(1979b), Burridge (1980) and Oden (1984). This topic is discussed in details in Bolduc(1985) and Cliff and Ord (1981).

Section A) ESTIMATION OF THE SAR(1) RESIDUAL

The spatial autoregressive of order 1 SAR(1) applied to the residual of a regression may be written using matrix algebra as:

(1) $y = X\beta + \epsilon$

(2) $\epsilon = \rho W\epsilon + u$

where u is a N-vector of i.i.d. random variables u_i , i=1,...,N, with mean 0 and variance σ^2, ρ is a spatial autocorrelation coefficient and W is the (NxN) matrix of the weights w_{ij} described in the introduction. X is a (NxK) fixed matrix of observations associated with the (Kx1) vector of unknown coefficient β. In this first part, assuming a W matrix fixed before estimation, we will consider different problems related to the estimation of the joint vector [β', σ^2 , ρ]. Noticing the similarity of the system formed by equation (1) and (2) with the usual (conditional on observation 1) temporal autocorrelation AR(1) formulation with W replaced by the matrix

$$C = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ . & . & \dots & . \\ 0 & 0 & . . & 1 \; 0 \end{bmatrix} ,$$

Hordijk (1974), to estimate β and ρ , suggested to use either the two step procedure of Durbin (1960) or the well known iterative method of Cochrane and Orcutt (1949). For temporal data, these methods are known to produce consistent and asymptotically efficient estimators. In these methods, ρ is estimated by Ordinary Least-Squares (OLS) with ϵ being replaced by the OLS estimate $\hat{\epsilon} = (y - X\hat{\beta})$.

Even if this extension seems quite natural, we will soon see that for spatial data, except for some specific configurations, the OLS estimation $\hat{\rho}$ of ρ in equation (2) is inconsistent. Before describing this result due to Whittle (1954) and Ord (1975), it is worth providing examples of some frequently encountered spatial dependency schemes. This exercise will help the reader to conceptualize how the different weighting matrices can be constructed to produce a formulation such as (1) and (2). Some technical details that will be used later such as eigenvalues of the matrix W are also given.

1) SOME RESIDUAL SPATIAL DEPENDENCY SCHEMES

1.1) The Unidimensional Bilateral Spatial Model: case 1

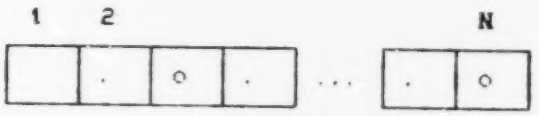
This model is written as:

$$(3)\quad y_i = \sum_{k=1}^{K} x_i^k \beta_k + \epsilon_i \quad , \qquad i = 1, \ldots, N$$

$$(4)\quad \epsilon_i = \frac{\rho}{2}(\epsilon_{i-1} + \epsilon_{i+1}) + u_i \qquad u_i \sim (0, \sigma^2) \ ,$$

where the division by 2 is justified by the fact that the model is then spatial stationary for $|\rho| < 1$.[1] The dependency pattern is depicted in figure 1.

FIGURE 1



Let C be the following contiguity matrix:

$$(5)\quad C = \begin{bmatrix} 0 & 1 & 0 & 0 & .. & 0 \\ 1 & 0 & 1 & 0 & .. & 0 \\ 0 & 1 & 0 & 1 & .. & 0 \\ . & . & . & . & . & 0 \\ 0 & 0 & 0 & . & . & 1 \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{bmatrix}$$

where, for $i \neq 1$ or N,

$$(6)\quad c_{ij} = \begin{cases} 1 & \text{if } i-1 = j \text{ or } i+1 = j \\ 0 & \text{otherwise .} \end{cases}$$

Putting $W = 1/2\ C$, when assuming $\epsilon_0 = \epsilon_{N+1} = 0$, equations (3) and (4) become :

$$(7)\quad y = X\beta + \epsilon$$

$$(8)\quad \epsilon = \rho W \epsilon + u \qquad u \sim (0, \sigma^2 I_N).$$

For this case, Ord (1975) obtains that $d_N = \det[\lambda I_N - W]$ defines a n^{th} order polynomial which satisfies the following

recurrence :

$$(9)\quad d_N = 2\lambda d_{N-1} - d_{N-2} .$$

As implied by theorem 1 which follows later, for the matrix W used in (8), the value of λ is such that $-1 \leq \lambda \leq 1$. Equation (9) then represents the recurrence relation of the Chebychev polynomial of the second type. Using this information, the N eigenvalues of W are:[2]

$$(10)\quad \lambda_i = \cos(\pi i/N+1) \qquad i = 1,\ldots,N$$

1.2) The quadrilateral AR of order 1 model QA(1):case 2

The stationary QA(1) for $|\rho| < 1$ is:

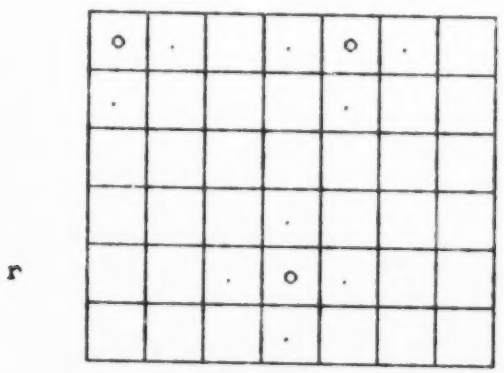
$$(11)\quad y_{r,s} = \sum_{k=1}^{K} x_{r,s}^{k} \beta_k + \xi_{r,s} , \qquad \begin{matrix} r=1,\ldots,P \\ s=1,\ldots,Q \end{matrix}$$

$$(12)\quad \xi_{r,s} = \frac{\rho}{2}(\xi_{r-1,s} + \xi_{r+1,s} + \xi_{r,s-1} + \xi_{r,s+1}) + \zeta_{r,s}$$

$$, \quad \zeta_{r,s} \sim (0, \sigma^2)$$

for which the dependency scheme is found in figure 2.

FIGURE 2



Let C_1 be the (PxP) matrix C of equation (5) and let C_2 be the same matrix but with dimension (QxQ). Equation (12), with the appropriate modifications for areas at the boundary of the region, may then be written as:

$$(13)\quad \xi = \frac{\rho}{4} (C_1 \xi + \xi C_2) + \zeta \quad ,$$

where ξ and ζ are matrices of dimension (PxQ). By stacking each

row in a (PQx1) column vector, or equivalently by exploiting the Kronecker product ⊗ properties, one can apply the vec formulation to finally obtain: [3]

$$(14)\quad \underset{(PQx1)}{vec(\xi)} = \frac{\rho}{4}\,[(C_1 \otimes I_Q) + (I_P \otimes C_2)]\,vec(\xi) + vec(\zeta).$$

Let $N = PQ$, $\epsilon = vec(\xi)$, $u = vec(\zeta)$ and

$$(15)\quad W = \frac{1}{4}\,[(C_1 \otimes I_Q) + (I_P \otimes C_2)].$$

By using the same type of transformation on equation (11), the QA(1) can be written as:

$$(16)\quad \begin{aligned} y_i &= \sum_{k=1}^{K} x_i^k \beta_k + \epsilon_i \\ \epsilon_i &= \rho \sum_{j=1}^{N} w_{ij}\,\epsilon_j + u_i \end{aligned} \qquad i = 1,\ldots,N$$

or using a matrix formulation:

$$(17)\quad \begin{aligned} y &= X\beta + \epsilon \\ \epsilon &= \rho W\epsilon + u \qquad u \sim (0, \sigma^2 I_N). \end{aligned}$$

For equation (15) Ord (1975) obtained for the eigenvalues $\lambda_{r,s}$ of W :

$$(18)\quad \lambda_{r,s} = \frac{1}{2}\,[\cos(r\pi/P+1) + \cos(s\pi/Q+1)]\,, \qquad \begin{aligned} r &= 1,\ldots,P \\ s &= 1,\ldots,Q\,. \end{aligned}$$

To find λ_i, $i = 1,\ldots,N$, one has to vectorize the matrix A of the $\lambda_{r,s}$'s computed in (18) as follows :

$$(19)\quad \underset{(Nx1)}{\lambda} = vec(A)\,.$$

1.3) The Spatial AR Wave Model of Order 1: case 3

This model implies for the residuals:

$$(20)\quad \xi_{r,s} = \frac{\rho}{2}\,(\xi_{r-1,s} + \xi_{r,s-1}) + \zeta_{r,s}\,,$$

with stationarity when $|\rho| < 1$. This kind of dependency is said to be unilateral since the dependency comes only from one side. A backward or forward formulation can be postulated

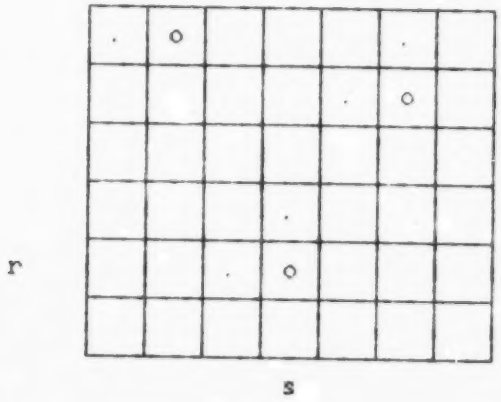
Using:

$$C_1 \underset{(P\times P)}{=} \begin{bmatrix} 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ . & . & \dots & . \\ 0 & 0 & .. & 1 \; 0 \end{bmatrix} \quad \text{and} \quad C_2 \underset{(Q\times Q)}{=} \begin{bmatrix} 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ . & . & \dots & . \\ 0 & 0 & .. & 1 \; 0 \end{bmatrix} ,$$

expression (20) with dependency pattern shown in figure 3 can be written as:

$$(21) \quad \xi = \underset{2}{\rho} (C_1 \xi + \xi C_2) + \zeta \; .$$

FIGURE 3



By vectorizing (21) as in section 1.2) one obtains:

$$(22) \quad \epsilon = \underset{2}{\rho} [(C_1 \otimes I_Q) + (I_P \otimes C_2')] \epsilon + u \; .$$

Under the form:

$$(23) \quad \begin{aligned} y &= X\beta + \epsilon \\ \epsilon &= \rho W \epsilon + u \; , \end{aligned}$$

the matrix W is such that:

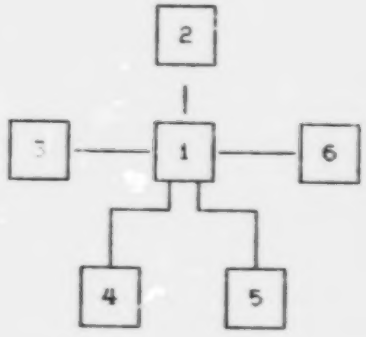
$$(24)\quad W_{(NxN)} = \begin{bmatrix} C_2' & O_Q & \cdots & O_Q & O_Q \\ I_Q & C_2' & \cdots & O_Q & O_Q \\ O_Q & \cdot & \cdot & \cdot & O_Q \\ \cdot & & \cdot & \cdot & \cdot \\ O_Q & \cdots\cdots & & I_Q & C_2' \end{bmatrix} .$$

The wave model formulation induces a lower triangular matrix for which all the terms on the diagonal are zero. The eigenvalues λ_i, $i = 1,\ldots,N$ of W are here all equal to zero.

1.4) Star Lattice Configuration: case 4

The two following examples are two extreme cases of contiguity . In figure 4, one can see that area 1 is dominant since each area is connected to area 1 without being mutually interconnected

FIGURE 4



Such a phenomenon, over N areas may be represented by the following structure of contiguity:

$$(25)\quad W_{(NxN)} = \begin{bmatrix} 1 & 1 & 1 & 1 & \ldots & 1 \\ 1 & 0 & 0 & 0 & \ldots & 0 \\ 1 & 0 & 0 & 0 & \ldots & 0 \\ . & . & . & . & \ldots & . \\ 1 & 0 & 0 & 0 & \ldots & 0 \\ 1 & 0 & 0 & 0 & \ldots & 0 \end{bmatrix} .$$

Using simple properties of the determinant of a matrix, the eigenvalues are easily shown to be : [4]

$$\lambda_i = 0 \qquad i = 1,\ldots,N-2$$

$$\lambda_{N-1} = -(N-1)^{1/2} \tag{26}$$

$$\lambda_N = (N-1)^{1/2} \quad .$$

1.5) Total Spatial Contiguity: case 5

The case where each area is related to every other area by the same amount $w_{ij} = k \quad \forall\, i \neq j$ leads to:

$$W = k \begin{bmatrix} 1 & 1 & 1 & 1 & \ldots & 1 \\ 1 & 1 & 1 & 1 & \ldots & 1 \\ 1 & 1 & 1 & 1 & \ldots & 1 \\ . & . & . & . & \ldots & . \\ 1 & 1 & 1 & 1 & & 1 \\ 1 & 1 & 1 & 1 & \ldots & 1 \end{bmatrix} \quad . \tag{27}$$

For such a matrix, the eigenvalues may be shown to be:

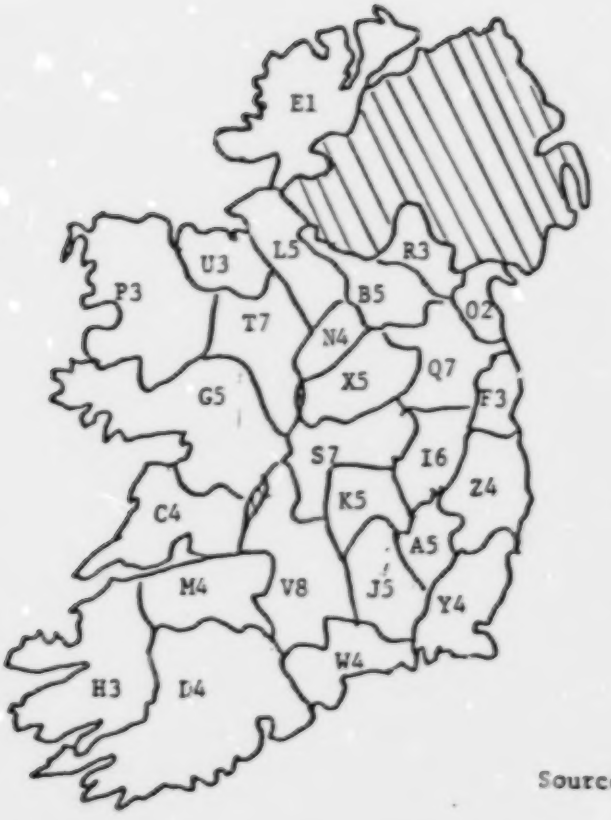
$$\lambda_1 = k.(N-1)$$

$$\lambda_i = -k \qquad i = 2,\ldots,N \tag{28}$$

1.6) Irregular Spatial Configurations : case 6

The last example depicted in figure 5 is described as irregular since compared to cases 1 to 5, each area has an unequal amount of neighbors of order 1. Two areas are called neighbors when they share a common border. An extension of this terminology to embed higher orders of contiguity may be found in Hordijk (1974) or Cliff and Ord (1981). In this paper only contiguity of order one is considered. To take into account the fact that for irregular lattices such as in figure 5, the number of neighbors, the size and the distances between areas are different, Cliff and Ord (1973) proposed to replace the boolean weighting formulation common to cases 1 to 5 by some functional forms w_{ij} (such as for example : $w_{ij}=d_{ij}^{-2}$ if i and j are contiguous, and 0 otherwise). Under irregular configurations, W has a form which is specific to the region under investigation. The eigenvalues for this case need to be evaluated with the aid of a computer.

Figure 5



Source : Geary (1968)

The stable SAR(1) model

For cases 1 to 3, which are possible realizations of some random field processes, the condition of stationarity was said to be $\rho < 1$. To obtain this result, it suffices to find the values of ρ for which the infinite moving average representation of the autoregressive process is convergent. As a natural extension of this result, Bodson and Peeters (1975) defined the SAR(1) of irregular spatial configurations, such as in case 6, as stationary when the values of ρ are such that the infinite sum:

$$\epsilon = (I-\rho W)^{-1}u = \sum_{i=0}^{\infty} \rho^i W^i u$$

is convergent. The values of ρ for which this condition holds may be shown to satisfy:

$$(29) \quad |\rho| < 1 \; / \; |\lambda_{MAX}(W)|$$

where $\lambda_{MAX}(W)$ is the maximal eigenvalue of matrix W.[5] Even though this definition seems quite natural, the term stationarity may become inappropriate for irregular configurations. As a matter of fact, the usual locational invariance properties of the ϵ_i process are violated in this case. Because of that, (29) will be referred here as the condition of stability of the SAR(1) process. In section 4, we will obtain for the stable SAR(1) formulation many desirable properties. In section 2, we describe the methods that has been suggested to estimate the parameters β and ρ of equations (1) and (2) for a W prespecified. The main result of that section leads to the conclusion that for general irregular spatial configurations, an OLS estimation of ρ is inconsistent.

2) OLS ESTIMATION OF ρ

The method of Durbin and Cochrane-Orcutt are based on an OLS estimation of ρ in:

$$(30) \quad \epsilon = \rho W\epsilon + u \qquad E(u)=0$$

$$E(uu')=\sigma^2 I_N,$$

which leads to:

$$(31) \quad \hat{\rho} = \epsilon'W\epsilon/\epsilon'W'W\epsilon = \rho + \epsilon'Wu/\epsilon'W'W\epsilon .$$

As opposed to the usual fixed regressor case for which the OLS solution is a linear sum of random variables, $\hat{\rho}$ in (31) is a non-linear function of the random vector ϵ. The common practice suggests to rely on the asymptotic properties of $\hat{\rho}$.[6] Since:

$$(32)\quad \operatorname*{plim}_{N\to\infty} \hat{\rho} = \rho + \frac{\operatorname*{plim}_{N\to\infty} \frac{\epsilon' W u}{N}}{\operatorname*{plim}_{N\to\infty} \frac{\epsilon' W' W \epsilon}{N}},$$

assuming that:

$$(33)\quad \operatorname*{plim}_{N\to\infty} \frac{\epsilon' W' W \epsilon}{N} = K < \infty,$$

if

$$(34)\quad \operatorname*{plim}_{N\to\infty} \frac{\epsilon' W u}{N} = 0,$$

$\hat{\rho}$ is a consistent estimator of ρ. Written differently, (34) is

$$(35)\quad \operatorname*{plim}_{N\to\infty} \frac{\sum_{i=1}^{N} u_i \left(\sum_{j=1}^{N} w_{ij} \epsilon_j \right)}{N}.$$

Then $\hat{\rho}$ is consistent when $\sum_{j=1}^{N} w_{ij} \epsilon_j$ (the spatial lag 1 of ϵ_i) is asymptotically independent of u_i. Note the similarity of this condition with the usual condition plim $\sum \epsilon_{t-1} u_t / T = 0$ in the temporal model $\epsilon_t = \rho \epsilon_{t-1} + u_t$. Sufficient conditions for (34) to be satisfied are:

$$(36)\quad \lim_{N\to\infty} E \left(\frac{\epsilon' W u}{N} \right) = 0,$$

and

$$(37)\quad \lim_{N\to\infty} \operatorname{Var} \left(\frac{\epsilon' W u}{N} \right) = 0.$$

Assuming here that (37) is fulfilled, we now study the behavior of (36). When the sum of the elements on the diagonal of the cross covariance matrix $E(We\ u')$, which is also:

$$(38)\quad E(W(I - \rho W)^{-1}uu') = \sigma^2 W(I - \rho W)^{-1},$$

is equal to zero, (36) is zero. When W is triangular, this condition is always satisfied. Indeed, for W triangular, say lower, $(I - \rho W)^{-1}$ is also lower triangular.

We just obtained that when the dependency is unidirectional (W triangular lower or upper) if the conditions (33) and (37) are met, the OLS estimation of ρ in (30) is consistent. Unfortunately, for a general W matrix (spatial model with multidirectional dependency), this condition is unlikely to be satisfied. Take for instance the case 2. Since:

$$(39)\quad \epsilon_{r,s} = \frac{\rho}{4}(\epsilon_{r-1,s} + \epsilon_{r+1,s} + \epsilon_{r,s-1} + \epsilon_{r,s+1}) + u_{r,s}$$

the fact that $\epsilon_{r-1,s} = f(\epsilon_{r,s})$ is sufficient to show that for that case, the assumption of independence between $u_{r,s}$ and the regressors is clearly violated. The similarity with the problem of bias in simultaneous equations is noteworthy. To summarize, except for some specific cases such as the wave model, the OLS estimation $\hat{\rho}$ of ρ is expected to be inconsistent. This very important result means that, if the dependency is anticipated to come from all directions (all cases except case 3) one should *not* use an OLS procedure to estimate ρ. It would be interesting to evaluate the intensity of this problem for some simple spatial structure. Of course no exact general results can be obtained since for irregular lattices the matrix W depends upon the geographical structure of the region. In the next section, we review the methods that have been suggested to produce a consistent estimation of ρ.

3) MAXIMUM LIKELIHOOD ESTIMATION OF β AND ρ

For spatial cases where the dependency is multidirectional, Ord (1975) and Hepple (1976), proposed to estimate the ρ of a SAR(1) residual formulation by a maximum likelihood (ML) methodology to produce a consistent estimation. As part of a more general formulation, the conditions under which the ML estimation of ρ is consistent will be investigated below. Ord (1975) suggested a modified Cochrane-Orcutt procedure. The estimation steps are essentially the same as those in time series except that the OLS estimation of ρ in equation (2) is replaced by a ML estimation based on the normality assumption. This procedure may easily be shown to be equivalent to the full maximum likelihood procedure of Hepple (1976) that we now describe.

3.1) Joint ML Estimation of β and ρ

Let $P = [I_N - \rho W]$ be a non singular (NxN) matrix such that equation (2) can be written as $P\epsilon = u$. Since the jacobian $J(u \to \epsilon) = |P|$ and since $J(\epsilon \to y) = 1$, under the following normality assumption $u \sim N(0, \sigma^2 I_N)$, the likelihood function of the SAR(1) residual model is: [7]

$$(40)\quad \mathcal{L}(\beta, \sigma^2, \rho) = (2\pi\sigma^2)^{-N/2} |P| \exp\left\{-\frac{(y-X\beta)'P'P(y-X\beta)}{2\sigma^2}\right\}$$

where $P = [I - \rho W]$ and $Var(\epsilon\epsilon') = \sigma^2 V = \sigma^2 P^{-1} P^{-1\prime}$ with $V^{-1} = P'P$.

Since P is function of ρ, for each different value of ρ, det P must be evaluated, where P is a (NxN) matrix. To overcome this difficulty when W is a matrix fixed before estimation, Ord (1975), using a well known property of the determinant of a matrix, suggested to replace det P by:

$$\det P = \det [I - \rho W] = \prod_{i=1}^{N} (1 - \rho\lambda_i)$$

where λ_i is the i^{th} eigenvalue of the matrix W. The fixity of W is such that for a given sample of N observations, det P changes only because of ρ. When replaced in the likelihood function, one obtains:

$$(41)\quad \mathcal{L}(\beta, \sigma^2, \rho) = (2\pi\sigma^2)^{-N/2} \prod_{i=1}^{N} (|1 - \rho\lambda_i|) \exp\{-\frac{(y-X\beta)'P'P(y-X\beta)}{2\sigma^2}\}$$

for which the log likelihood is:

$$(42)\quad \ell(\beta, \sigma^2, \rho) = -\frac{N}{2} \ln 2\pi - \frac{N}{2} \ln \sigma^2 + \sum_{i=1}^{N} \ln|(1 - \rho\lambda_i)| - \frac{(y - X\beta)'P'P(y - X\beta)}{2\sigma^2} .$$

The first order conditions lead to the usual expressions $\tilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$ and $\tilde{\sigma}^2 = (y - X\tilde{\beta})'V^{-1}(y - X\tilde{\beta})/N$, both of which are functions of ρ since $V^{-1} = P'P$. Replacing β and σ^2 by these formulas in (41) leads to a doubly concentrated likelihood function that depends only upon ρ. The problem then becomes equivalent to minimizing:

$$(43)\quad \mathcal{L}^{**}(\rho) \propto |\prod_{1}^{N}(1 - \rho\lambda_1)|^{-2/N} \tilde{\sigma}^2$$

or the natural log:

$$(44)\quad \ell^{**}(\rho) \propto \ln \tilde{\sigma}^2 - \frac{2}{N} \sum_{1}^{N} \ln|(1 - \rho\lambda_1)| ,$$

where $\tilde{\sigma}^2 = \frac{(y - X\tilde{\beta})'V^{-1}(y - X\tilde{\beta})}{N}$.

Expression (43) brings out, as in time series, the difference between the ML estimation and the methods such as Cochran-Orcutt, Durbin two-step or non-linear least-squares that aims to minimize the χ^2 form, $\tilde{\sigma}^2 = (y - X\tilde{\beta})'V^{-1}(y - X\tilde{\beta})/N$. For a temporal AR(1) residual model, the ML estimation is intended to minimize :

$$(45)\quad \mathcal{L}^{AA}(\rho) \propto (1 - \rho^2)^{-1/N} \frac{(y - X\tilde{\beta})'V^{-1}(y - X\tilde{\beta})}{N} .$$

When $|\rho|$ is not too close to 1, $(1 - \rho^2)^{-1/N}$ becomes negligible when $N \to \infty$ such that in that case:

$$(46)\quad \lim (1-\rho^2)^{-1/N} = 1 ,$$

which leads to the asymptotic equivalency of the method of estimation . The SAR(1) equivalent of (45), as seen in (43) is:

$$(47)\quad \ell^{**}(\rho) \propto |\Pi_{1}^{N}(1 - \rho\lambda_{1})|^{-2/N} \frac{(y - X\tilde{\beta})'V^{-1}(y - X\tilde{\beta})}{N}$$

where $|\Pi_{1}^{N}(1 - \rho\lambda_{1})| = J(u \to \epsilon)$ that we now note J.

Griffith (1980) was able to prove that as $N \to \infty$, $\lim J^{-2/N} = 1$, probably by neglecting the fact that as $N \to \infty$ the dimension of W and thus the number of terms and the values of λ_1 change. This result led him to conclude that the asymptotic equivalence of the time series methodologies applies also for the spatial AR(1) case. As it should be expected from the results of section 2, when the spatial dependency is multidirectional this limit is not equal to 1. Ord (1981) provides some simple examples that show that in general the result of Griffith is false. Another result obtained by Whittle (1954) shows that for the QA(1) described in section 1 "case 2",

$$\lim_{N\to\infty} J^{-2/N} = \exp\left[\sum_{j=1}^{\infty} \frac{1}{j} \binom{2j}{j} \rho^{2j}\right] .$$

which is definitely different from 1. This of course constitutes another indication that in the case of multidirectional dependency, the OLS estimation of ρ is inconsistent.

These last results imply that the jacobian in the likelihood function of the SAR(1) model is a term that, except for some specific cases already discussed, never vanishes even asymptotically. In the following section, we present the analytical behavior of the concentrated likelihood function:

$$(48)\quad \ell^{**}(\rho) \propto +\ln|P| - \frac{N}{2} \ln \tilde{\sigma}^2,$$

where $\tilde{\sigma}^2 = \frac{(y - X\tilde{\beta})'V^{-1}(y - X\tilde{\beta})}{N}$.

It is obvious that maximizing (48) is the same as minimizing (44). We will see that due to the presence of the jacobian term, the behavior of the likelihood function of the SAR(1) may sometimes be quite erratic.

4) ANALYTICAL BEHAVIOR OF THE SAR(1)'S LIKELIHOOD FUNCTION

Let W be the contiguity matrix of any SAR(1) process. Since each element of the diagonal of W is zero, tr(W) = 0 so that the sign of W is always indeterminate. In other words, there is always negative and positive eigenvalues λ_i . When we order them in the increasing order, we have:

$$\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_S < 0 < \lambda_{S+1} \leq \lambda_{S+2} \leq \ldots \leq \lambda_N$$

meaning that λ_S is the largest negative eigenvalue and λ_{S+1} the smallest positive eigenvalue. We then can write:

$$(49)\quad \lambda_S^{-1} \leq \lambda_{S-1}^{-1} \leq \ldots \leq \lambda_1^{-1} < 0 < \lambda_N^{-1} \leq \ldots \leq \lambda_{S+1}^{-1}$$

where since tr W = 0, $\sum_{i=1}^{N} \lambda_i = 0$

Expression (41) allows us to write (48) as:

$$(50)\quad \mathcal{L}^{**}(\rho) \propto \sum_{i=1}^{N} \ln|(1 - \rho\lambda_i)| - \frac{N}{2} \ln \tilde{\sigma}^2 .$$

The equation (49) defines for the likelihood function (50), at most N indetermination points [Ord(1981)]. Indeed, for each value where $\rho = 1/\lambda_i$, $\ln(0) = -\infty$. The ML solution then must be located on one of the segments determined by (49). Moreover, by (29), the SAR(1) is stable when $|\rho| < 1/|\lambda_{MAX}(W)|$. This means that to produce a stable solution, ρ must belong to one of the two following open interval:

$$(51)\quad \left.\right]\!\!-\!\!-\!\!-

The knowledge of those results can now help us to describe the behavior of the function $\ln|P|$ in ℓ^{**}. The first and second derivatives of $\ln|P|$ with respect to ρ give:

$$(53) \qquad \frac{\partial \sum_{i=1}^{N} \ln|(1-\rho\lambda_i)|}{\partial\rho} = -\sum_{i=1}^{N} \frac{\lambda_i}{(1-\rho\lambda_i)}$$

and

$$(54) \qquad \frac{\partial^2 \ln|P|}{\partial\rho^2} = -\sum_{i=1}^{N} \frac{\lambda_i^2}{(1-\rho\lambda_i)^2} < 0.$$

Since for any value of ρ different from one of the $1/\lambda_i$'s, equation (54) is negative, on each open segment between successive $1/\lambda_i$'s $\ln|P|$ is strictly concave. For each value of $\rho = 1/\lambda_i \ \forall\ i$, $|P| = 0$, such that $\ln|P| = -\infty$. A further result can be obtained on the segment $1/\lambda_1 < 0 < 1/\lambda_N$. Since (53) is zero when $\rho = 0$, [$\Sigma\lambda_i = 0$ always] the maximum of the concave function $\ln|P|$ defined in this range is found at $\rho = 0$. At that value, $|P| = 1$ so that $\ln|P| = 0$. As expected, in the absence of spatial autocorrelation, the jacobian term disappears. We now know the behavior of one of the two terms that form ℓ^{**}. Since $-N/2\ \ln\tilde{\sigma}^2$ is usually a smooth continuous function, ℓ^{**} will mostly follow the behavior of $\ln|P|$. The highly negative value of $\ln|P|$ at each $\rho = 1/\lambda_i$ and the strict concavity on each segment defined by (49) will then be reflected in ℓ^{**}. To give an example, a SAR(1) model estimated in section C for which the matrix W is the (26x26) boolean "0 or 1" first order of contiguity matrix associated with figure 5, leads to the likelihood function depicted in figure 6 for the interval $-2 \leq \rho \leq 2$. The decomposition in terms of $\ln|P|$ and $-N/2\ \ln\sigma^2$ on the $0 \leq \rho < 1$ domain is presented in figures 7 and 8 respectively.

4.1) Implication of a Normalized Contiguity Matrix

We now consider two type of contiguity matrices that are such that the corresponding SAR(1) model is stable for $-1 < \rho < 1$. These normalizations are such that the likelihood function on the stable domain of ρ, is expected to be strictly concave with respect to ρ. Before describing these matrices, we need the result of the next theorem:

Figure 6

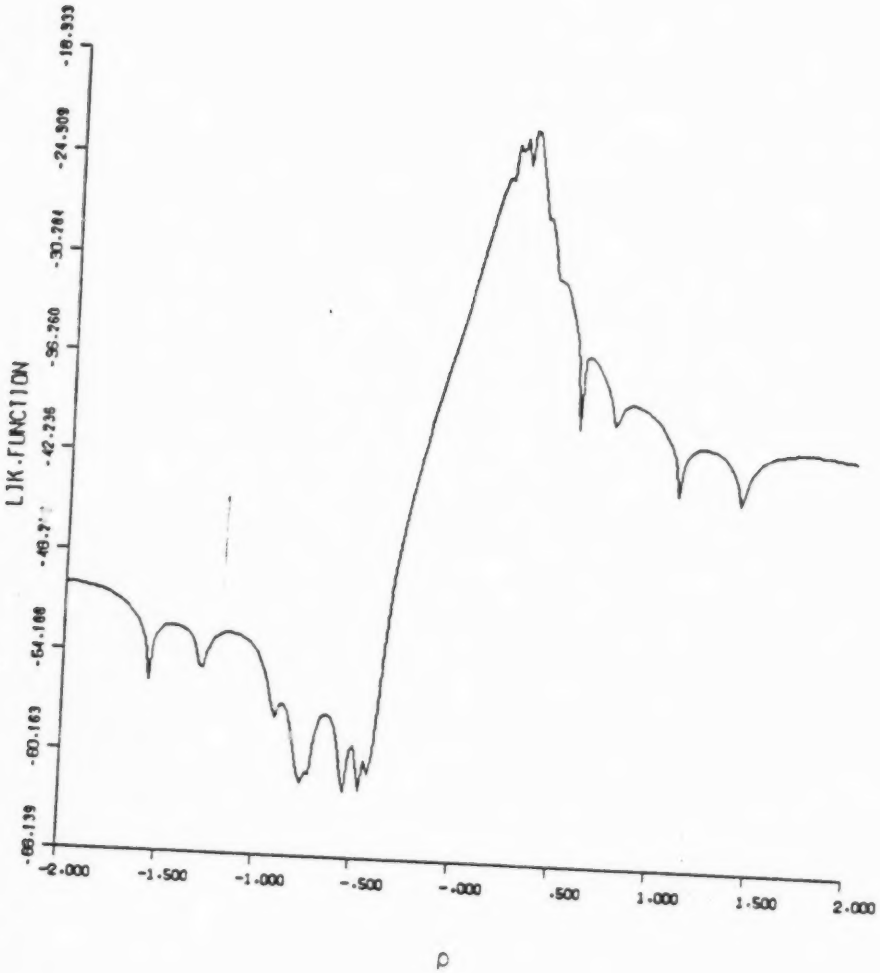

LIK.FUNCTION
-18.933
-24.909
-30.784
-36.760
-42.736
-48.7
-54.188
-60.163
-66.139
-2.000
-1.500
-1.000
-.500
-.000
.500
1.000
1.500
2.000
ρ

Figure 7

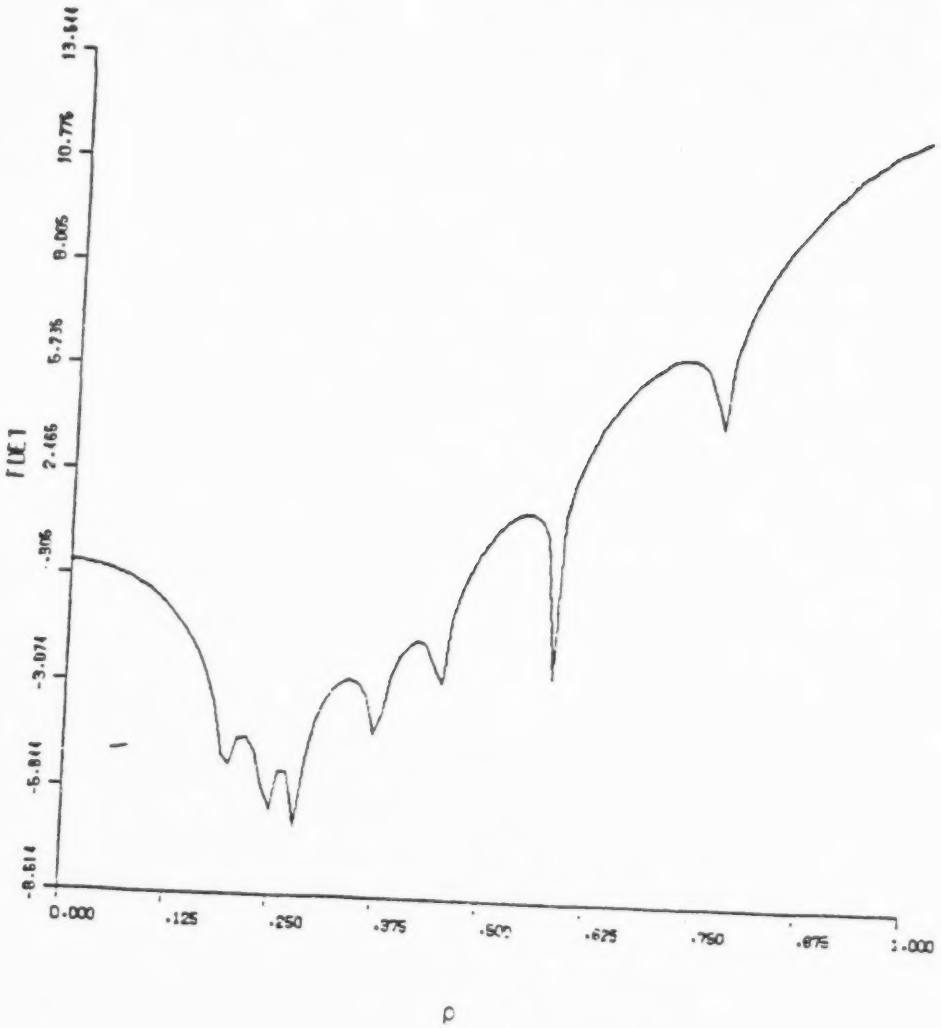
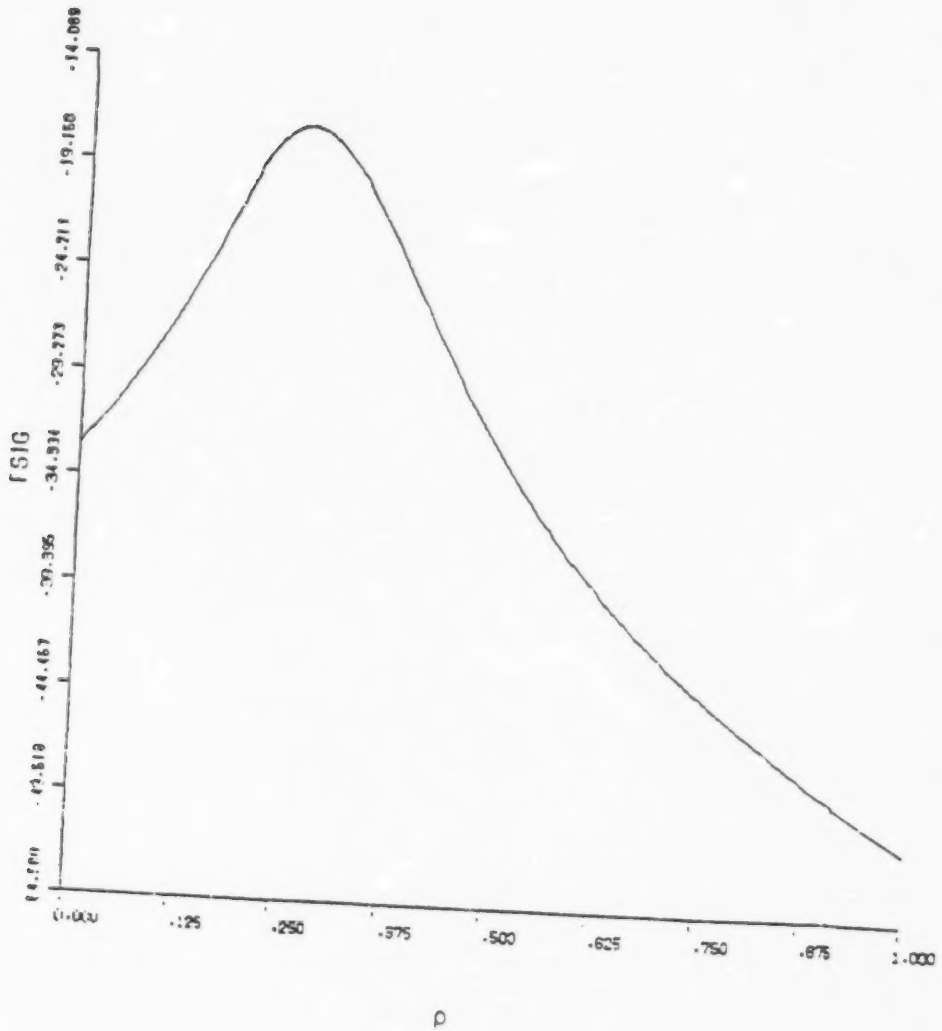

13.644
10.776
8.005
5.235
2.465
-.306
-3.076
-5.846
-8.616
F(E)
0.000
.125
.250
.375
.500
.625
.750
.875
1.000
1
p

Figure 8


FSIG
-14.068
-19.158
-24.211
-29.273
-34.334
-39.396
-44.457
-49.518
0.000
.125
.250
.375
.500
.625
.750
.875
1.000
p

THEOREM 1

Let λ be an eigenvalue of a matrix A, then

$$|\lambda| \leq \sum_j |a_{ij}|$$

Proof:

$(A - \lambda I)x = 0$ can be written as: $\sum_{j=1}^{N} a_{ij} x_j = \lambda x_i, \ i = 1, 2, \ldots, N$

$$\forall i \quad |\lambda| |x_i| = |\sum a_{ij} x_j| \leq \sum |a_{ij}| |x_j| \leq \sum |a_{ij}| \max_j |x_j|$$

$$\forall i \quad |\lambda| \max_i |x_i| \leq \sum_j |a_{ij}| \max_j |x_j|$$

so that $|\lambda| \leq \sum_j |a_{ij}|$.

Non symmetrical contiguity matrix M

The first normalized contiguity matrix that we consider is often noted in the literature as M. M is such that:

$$(55) \quad m_{ij} = \frac{w_{ij}}{\sum_{j=1}^{N} w_{ij}}$$

where $m_{ii} = 0 \ \forall i$ and $\sum_j^N m_{ij} = 1 \ \forall i$.

Even if W is symmetrical, this normalization makes M non symmetric. Let λ_i be the eigenvalues of matrix M. Since the w_{ij}'s are non-negative terms, the last theorem implies that $\lambda \leq +1$. Let ι be a (Nx1) vector of 1. The condition $\sum m_{ij} = 1$ which is also in matrix notation $M\iota = \iota$, shows that +1 is the eigenvalue of the matrix M associated with the eigenvector ι. These two last facts imply: $\lambda_{MAX}(M) = +1$. So the model:

$$(56) \quad y = X\beta + \epsilon$$

$$(57) \quad \epsilon = \rho M\epsilon + u \qquad u \sim N(0, \sigma^2 I_N)$$

is stable when $|\rho| < 1$.

By applying the results of section 4, one may deduce the behavior of $\ln|P|$ over the stable range of ρ depicted in figure 9. This suggest the presence of a strictly concave continuous likelihood function with respect to ρ on the domain $-1 < \rho < 1$. This will be proved to hold asymptotically later on. In the example that led to figure 6, when we replace the matrix W by the normalized matrix M, the likelihood function of figure 10 is produced. The nice behavior of ℓ^{**} on the domain $|\rho| < 1$ is quite remarkable.

FIGURE 9

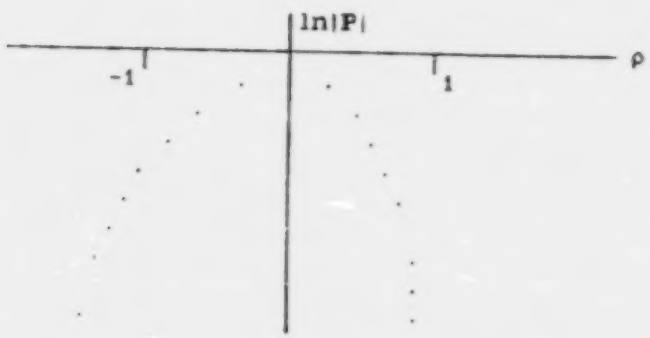
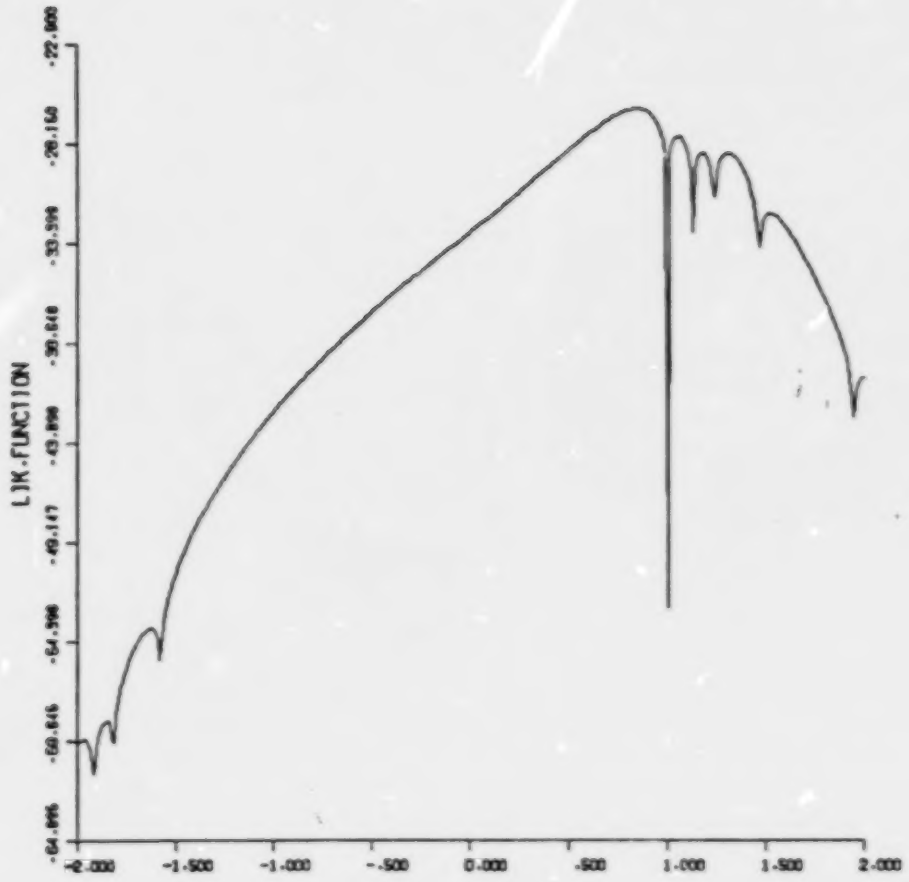


Figure 10


LIK.FUNCTION
-22.000
-28.154
-33.308
-38.462
-43.615
-48.769
-53.923
-59.077
-64.231
-2.000
-1.500
-1.000
-.500
0.000
.500
1.000
1.500
2.000
ρ

Symmetrical contiguity matrix $\underset{\sim}{W}$.

Taking a symmetric matrix W, we will now define a type of normalization such that the produced matrix is symmetrical and also such that $|\rho| < 1$ define a stable SAR(1) model. Let:

$$(58)\quad \underset{\sim}{w}_{ij} = \frac{w_{ij}}{\left(\sum_{i}^{N} w_{ij}\right)^{1/2} \left(\sum_{j}^{N} w_{ij}\right)^{1/2}} .$$

To show that the maximal eigenvalue is +1, we only have to use the fact that the non-zero eigenvalues of the product of matrices does not depend on the order of the matrices in the product. Let D be a diagonal matrix with elements: $d_{ii} = 1/\sum_{j} w_{ij}$. We then have:

$$(59)\quad \begin{aligned} M &= W\,D \\ \underset{\sim}{W} &= D^{1/2}\,W\,D^{1/2}. \end{aligned}$$

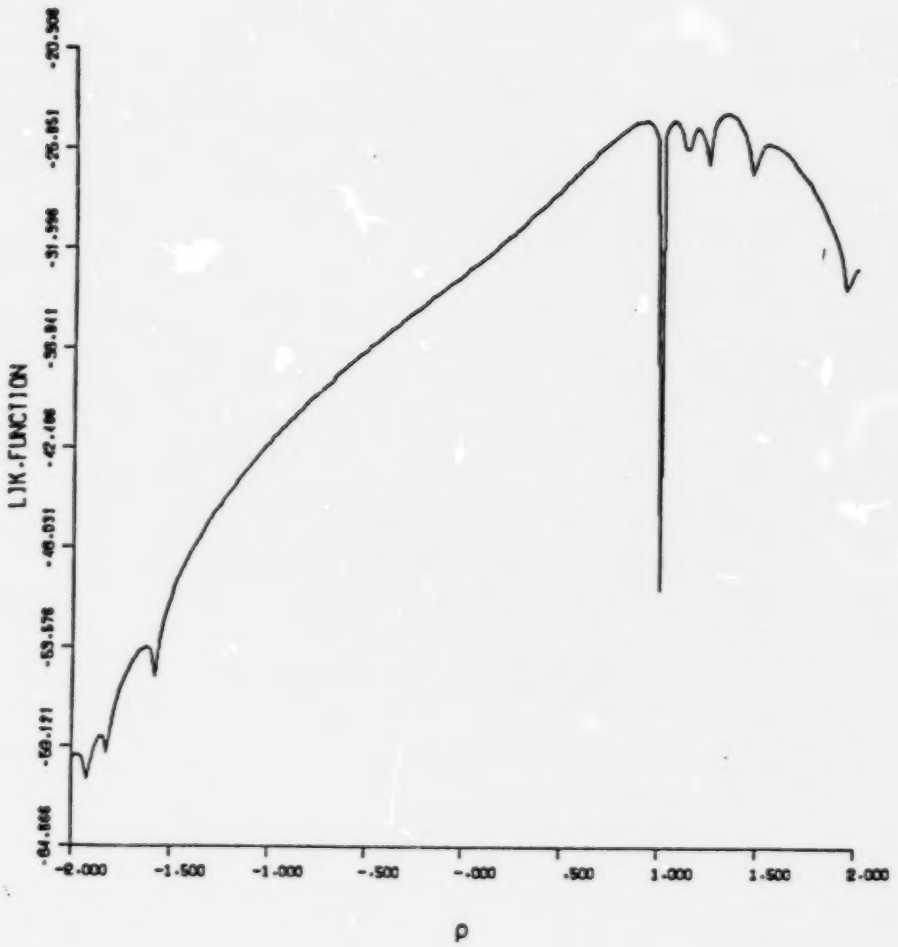
Using this last property of the determinant we have:

$$\lambda(W\,D) = \lambda(W\,D^{1/2}\,D^{1/2}) = \lambda(D^{1/2}\,W\,D^{1/2}) = \lambda(\underset{\sim}{W})$$

which proves that for the symmetrical contiguity matrix case, the solution is stable when $|\rho| < 1$ is satisfied. Our likelihood function for this case is depicted in figure 11. Because those two normalized matrices have the same eigenvalues, figures 10 and 11 have exactly the same points of indetermination.

We can note in figures 6 and 11 the presence of a global maximum outside the stable solution range $|\rho| < 1$. However in figure 10, the global solution defines a stable model. This last phenomenon raises two questions: 1) Is the presence of a global maximum outside the stable region related to the fact that for these cases, the contiguity matrix is symmetric? 2) Do we restrict the search of a solution to the well behaved portion of the likelihood function [stable solution] or do we also look for a "better" non-stable solution [global solution]. To answer the first question, further applications are needed. The second, which is of more interest, depends on the view of the researcher. In our opinion, the stable solution is in many respects more desirable. As we will formally show in section 5 that follows, if the true value ρ_o belongs to the stable range of ρ, the ML procedure leads to globally consistent and asymptotically normal estimators of all the unknown parameters of the generalized SAR(1) model.

Figure 11


LIK-FUNCTION
-20.506
-26.851
-31.596
-38.841
-42.186
-46.031
-53.576
-58.121
-64.866
-2.000
-1.500
-1.000
-.500
-.000
.500
1.000
1.500
2.000
ρ

The main conclusion of this section is the following:

If the SAR(1) is postulated to be stable by assuming one of the two discussed normalized form for the contiguity matrix, the behavior of the likelihood function with respect to ρ on the unit interval is extremely smooth. Of course this conclusion does not change when one relaxes the fixity of the w_{ij}'s by postulating an estimable function w_{ij} such as described in the introduction. In other words, by assuming a normalized contiguity matrix, one can undertake the estimation of the vector θ along with ρ without worrying with computational problems due to multiple extrema. In section C, we provide some convincing examples that show the almost insurmountability of the estimation problems caused by an unormalized contiguity matrix.

5) THE ASYMPTOTIC PROPERTIES OF THE SAR(1) RESIDUAL

The asymptotic properties of the maximum likelihood estimation of the SAR(1) model, even with W fixed has never been investigated before. That we will do now. In section 1, we will derive some results with W assumed fixed. In section 2, the case with estimable W will be considered. Details concerning the approach used in this section may be found in Amemiya (1985), Burguette Gallant and Souza (1982), Gourieroux Monfort and Trogon (1984) or White (1982).

Remarks:

1) In section B, the ML estimations for β and σ^2 are shown to be: $\tilde{\beta}=(X'\tilde{V}^{-1}X)^{-1}X'\tilde{V}^{-1}y$ $\tilde{\sigma}^2=(y-X\tilde{\beta})'\tilde{V}^{-1}(y-X\tilde{\beta})/N$.
The properties of consistency and of asymptotic normality of these estimators depend directly on the consistency properties of the estimators that form $\tilde{V}$. Because the likelihood function can be concentrated with respect to β and σ^2, the asymptotic properties of β, σ^2 and ρ depend on the properties of ρ in the following model where ϵ is assumed observed.

$$(60)\quad \epsilon = \rho W\epsilon + u \qquad u \sim N(0, \sigma^2 I_N).$$

where ρ maximizes the following function:

$$(61)\quad \ell_N(\sigma^2, \rho) = -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln\sigma^2 + \ln|P| - \frac{\epsilon P'P\epsilon}{2\sigma^2}.$$

2) To simplify, the matrix W will designate either the non-symmetrical normalized matrix M or the symmetrical normalized one. For these cases, the model is stable when $|\rho|<1$. Since the random variables ϵ_i in (60) are neither independent nor identically distributed, we will need to use for our asymptotic proofs a general theorem such as theorem 4.1.1 of Amemiya(1985). We will use this theorem to prove the consistency of the ML estimation of ρ in model (60) where W is assumed fixed.

5. 1) Consistency of ML Estimation, Matrix W Fixed

Make the assumptions

A) ρ_o belongs to a compact subspace $\Gamma = [-(1-\delta), (1-\delta)]$ of $|\rho|<1$ were δ is a very small positive number.
B) ℓ_N is continuous in $\rho\epsilon\Gamma$ for all ϵ and is a measurable function of ϵ for all $\rho\epsilon\Gamma$.
C) $N^{-1}\ell_N$ converges to a non stochastic function $Q(\rho)$ in probability uniformly in $\rho\epsilon\Gamma$ as $N\to\infty$ and $Q(\rho)$ attains a unique global maximum at ρ_o.

Then $\hat{\rho}_N$ that maximizes (61) converges to ρ_o in probability.

Proof: Amemiya (1985) p. 107.

If assumptions A) to C) are satisfied, $\hat{\rho}_N$ is said to be a global consistent estimator for ρ_o. Since the true model is:

$$(62)\quad \epsilon = \rho_o W\epsilon + u \qquad u \sim N(0, \sigma_o^2 I_N),$$

replacing ϵ in (61) by $[I-\rho_o W]^{-1}u$, which we note $P_o^{-1}u$, we can write:

$$(63)\quad \frac{1}{N}\ell_N = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 + \frac{1}{N}\ln|P| - \frac{u'[I-\rho_o W]^{-1\prime}P'P[I-\rho_o W]^{-1}u}{2N\sigma^2}.$$

Since $P = [I-\rho W] = [(I-\rho_o W)-(\rho-\rho_o)W]$, $PP_o^{-1} = [I-(\rho-\rho_o)W(I-\rho_o W)^{-1}]$,

(63) can equivalently be written:

$$(64)\quad \frac{1}{N}\ell_N = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 + \frac{1}{N}\ln|(I-\rho_o W)-(\rho-\rho_o)W|$$

$$-\frac{1}{2\sigma^2}\left\{\frac{u'u}{N} - \frac{2}{N}(\rho-\rho_o)\epsilon'W'u + \frac{1}{N}(\rho-\rho_o)^2\epsilon'W'W\epsilon\right\}$$

By Kolmogorov LLN 2, $\operatorname{plim}_{N\to\infty} \frac{u'u}{N} = \sigma_o^2$.

Assume that the following limits in probability relations hold:

$$\operatorname*{plim}_{N\to\infty} \frac{\epsilon'W'u}{N} = \sigma_o^2 \lim_{N\to\infty} \frac{1}{N} \operatorname{tr}[(I-\rho_o W)^{-1\prime}W'] = \sigma_o^2 m_1$$

$$\operatorname*{plim}_{N\to\infty} \frac{\epsilon'W'W\epsilon}{N} = \sigma_o^2 \lim_{N\to\infty} \frac{1}{N} \operatorname{tr}[(I-\rho_o W)^{-1\prime}W'W(I-\rho_o W)^{-1}] = \sigma_o^2 m_2$$

and;

$$\lim_{N\to\infty} \frac{1}{N} \ln|P| = m_0 < \infty .$$

For $\rho, \rho_o \in \Gamma$, recalling the uniform behavior caused by the normalized contiguity matrices, this convergence is clearly uniform. We then have:

$$(65)\quad Q(\rho) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 + \lim_{N\to\infty} \frac{1}{N} \ln|(I-\rho_o W)-(\rho-\rho_o)W|$$

$$-\frac{1}{2\sigma^2}\sigma_o^2 + (\rho-\rho_o)\frac{1}{\sigma^2}\sigma_o^2 m_1 - (\rho-\rho_o)^2 \frac{1}{2\sigma^2}\sigma_o^2 m_2 .$$

When $N^{-1}\ell_N$ converges uniformly to the well defined limit $Q(\rho)$ implied by the previous conditions, $\hat{\rho}$ is a global consistent estimator if we can show that the maximum of $Q(\rho)$ with respect to ρ is unique. We have the following proposition:

proposition 1

On the stable interval $-1<\rho<1$, when one of the two considered normalizations is applied, if ρ_o belongs to the interval, then $Q(\rho)$ is uniquely maximized at $\rho=\rho_o$.

proof

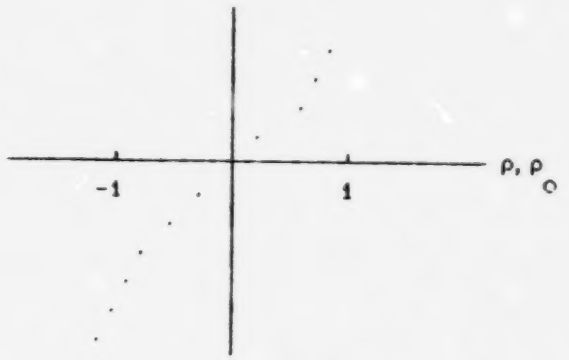
Similarly to (53) and (54), when we take the derivative of $Q(\rho)$ with respect to $(\rho-\rho_o)$ we can write, since $\sigma^2=\sigma_o^2$ uniquely :

$$(66)\quad \frac{\partial\, Q(\rho)}{\partial(\rho-\rho_o)} = -\lim_{N\to\infty} \frac{1}{N} \sum_{i}^{N} \frac{\lambda_i}{(1-(\rho-\rho_o)\lambda_i-\rho_o\lambda_i)} + \lim_{N\to\infty} \frac{1}{N} \sum_{i}^{N} \frac{\lambda_i}{(1-\rho_o\lambda_i)}$$

$$-(\rho-\rho_o)\lim_{N\to\infty} \frac{1}{N} \sum_{i}^{N} \frac{\lambda_i^2}{(1-\rho_o\lambda_i)^2} .$$

As expected (66) is zero when $\rho=\rho_o$. Call $(\rho-\rho_o)$: τ. We will now prove that for the stable SAR(1) model, (66) cannot be 0 when $\tau\neq 0$. The development below (54) and figure 9 imply for $\sum \frac{\lambda_i}{i(1-\rho\lambda_i)}$ [negative of equation (53)] of a stable SAR(1) model, the pattern depicted in figure 12.

FIGURE 12



Assume that for $\tau\neq 0$ (66) can be zero. This implies that the next implicit equation can hold

$$(67)\ \tau = \left[\lim_{N\to\infty} \frac{1}{N} \sum \frac{\lambda_i^2}{i(1-\rho_o\lambda_i)^2} \right]^{-1} \left[-\lim_{N\to\infty} \frac{1}{N} \sum \frac{\lambda_i}{i(1-\rho\lambda_i)} + \lim_{N\to\infty} \frac{1}{N} \sum \frac{\lambda_i}{i(1-\rho_o\lambda_i)} \right]$$

For $0<\rho_o<1$, figure 12 implies:

$$\lim_{N\to\infty} \frac{1}{N} \sum \frac{\lambda_i}{i(1-\rho_o\lambda_i)} < \lim_{N\to\infty} \frac{1}{N} \sum \frac{\lambda_i}{i(1-\rho\lambda_i)} \quad \text{for all } \rho > \rho_o \text{ or } \tau>0$$

$$\lim_{N\to\infty} \frac{1}{N} \sum \frac{\lambda_i}{i(1-\rho_o\lambda_i)} > \lim_{N\to\infty} \frac{1}{N} \sum \frac{\lambda_i}{i(1-\rho\lambda_i)} \quad \text{for all } \rho < \rho_o \text{ or } \tau<0.$$

Applying these relations to (67), since the term inverted there is positive, a contradiction occurs unless $\tau=0$. Doing the same for the negative part $-1<\rho_o<0$ leads to the same conclusion. We can thus conclude that $Q(\rho)$ is uniquely maximized at ρ_o when the parameter space is constrained to be a compact subspace of the stable interval.

As a corollary, if the previous assumptions hold, we obtain that the maximum likelihood estimation of the joint vector $\eta=[\beta', \sigma^2, \rho]'$ is globally consistent when the SAR(1) with appropriate normalized contiguity matrix is stable. Of course ρ_o must belong to the stable interval. When ρ_o lies outside this interval, the consistency is local. In extending this result to the estimable W matrix case, we now use in the next section this last concept of consistenc.

5.2) Consistency of ML estimation, matrix W estimated

The elements of W are non-linear functions of the parameter vector θ. Since in our proof, we will be using a linear approximation around the true valued vector $\gamma_o=[\rho_o, \theta_o']'$, we will obtain the conditions under which in an open neighborhood of γ_o, a root $\hat{\gamma}$ of $\partial\ell/\partial\gamma=0$ is a local consistent estimator of γ_o. To simplify the notation, we will assume here that θ is a scalar. Using a more tedious notation, the generalization to a H-vector θ of the following results will be shown to hold. More details is given below. To obtain this result, we now use theorem 4.1.2 of Amemiya (1985).

Make the assumptions

A) $\gamma=[\rho, \theta]$ belongs to an open subset Γ of the euclidean 2-space, with $\gamma_o \in \text{int}(\Gamma)$.
B) $\ell_N(\epsilon, \gamma)$ in (61) is a measurable function of ϵ for all $\gamma\in\Gamma$ and $\partial\ell_N/\partial\gamma$ exist and is continuous in an open neighborhood $N_1(\gamma_o)$ of γ_o. (This implies that $Q(\gamma)$ is continuous for $\gamma\in N_1$).
C) There exist an open neighborhood $N_2(\gamma_o)$ of γ_o such that $N^{-1}\ell_N$ converges to a non stochastic function $Q(\gamma)$ in probability uniformly in γ in $N_2(\gamma_o)$ and $Q(\gamma)$ attains a strict local maximum at γ_o.

If these assumptions are satisfied, there exist a consistent root of the equation $\partial\ell_N/\partial\gamma=0$, corresponding to a local maxima.

Proof: Amemiya (1985) p. 111

For the same reasons as in section 5.1, we will study the estimation properties of ρ and θ in model (60) and (61) where ϵ is assumed observed.

Taking a first order approximation of $P=[I-\rho W(\theta)]$, we have:

$$P = [\ (I-\rho_o W_o) - (\rho-\rho_o)W_* - (\theta-\theta_o)\rho_* G_*\ ],$$

where $\partial W/\partial\theta$ is denoted G and the * subscript means the evaluation at γ^* which lies between γ and γ_o. Since the true model is:

(68) $\epsilon = (I-\rho_o W_o)^{-1}u = P_o^{-1}u,$

the following relation between P and P_o holds:

$$PP_o^{-1} = [I - (\rho-\rho_o)W_N P_o^{-1} - (\theta-\theta_o)\rho_N G_N P_o^{-1}].$$

Since we will need to evaluate $\epsilon'P'P\epsilon$ and since $\epsilon = P_o^{-1}u$, we have:

$$\begin{aligned}(69)\ \epsilon'P'P\epsilon = {} & u'u - 2(\rho-\rho_o)u'W_N\epsilon - 2(\theta-\theta_o)\rho_N u'G_N\epsilon \\ & + 2(\rho-\rho_o)(\theta-\theta_o)\rho_N\epsilon'W_N'G_N\epsilon + (\rho-\rho_o)^2\epsilon'W_N'W_N\epsilon \\ & + (\theta-\theta_o)^2\rho_N^2\epsilon'G_N'G_N\epsilon.\end{aligned}$$

As an extension of the conditions assumed in the previous section, we assume that the following terms converge uniformly in probability to the listed finite limits:

$$\operatorname*{plim}_{N\to\infty}\frac{\epsilon'W_N'u}{N} = \sigma_o^2 \lim_{N\to\infty}\frac{1}{N}\operatorname{tr}[(I-\rho_o W)^{-1}W_N'] = \sigma_o^2 m_{11}$$

$$\operatorname*{plim}_{N\to\infty}\frac{\epsilon'G_N'u}{N} = \sigma_o^2 \lim_{N\to\infty}\frac{1}{N}\operatorname{tr}[(I-\rho_o W)^{-1}G_N'] = \sigma_o^2 m_{12}$$

$$\operatorname*{plim}_{N\to\infty}\frac{\epsilon'W_N'G_N\epsilon}{N} = \sigma_o^2 \lim_{N\to\infty}\frac{1}{N}\operatorname{tr}[(I-\rho_o W)^{-1}W_N'G_N(I-\rho_o W)^{-1}] = \sigma_o^2 m_{21}$$

$$\operatorname*{plim}_{N\to\infty}\frac{\epsilon'W_N'W_N\epsilon}{N} = \sigma_o^2 \lim_{N\to\infty}\frac{1}{N}\operatorname{tr}[(I-\rho_o W)^{-1}W_N'W_N(I-\rho_o W)^{-1}] = \sigma_o^2 m_{22}$$

$$\operatorname*{plim}_{N\to\infty}\frac{\epsilon'G_N'G_N\epsilon}{N} = \sigma_o^2 \lim_{N\to\infty}\frac{1}{N}\operatorname{tr}[(I-\rho_o W)^{-1}G_N'G_N(I-\rho_o W)^{-1}] = \sigma_o^2 m_{23}$$

and;

$$\lim_{N\to\infty}\frac{1}{N}\ln|P| = m_o < \infty .$$

For judicious choices of functional forms w_{ij} included in a *stable* SAR(1) model with normalized contiguity matrix, these conditions are likely to be satisfied. Examples of possible forms are given in section C. As for (65), when evaluated at $\sigma^2=\sigma_o^2$, we end up with the following limit function:

$$\begin{aligned}(70)\ Q(y) \propto {} & \lim_{N\to\infty}\frac{1}{N}\ln|P| + (\rho-\rho_o)m_{11} + (\theta-\theta_o)\rho_N m_{12} \\ & -(\rho-\rho_o)(\theta-\theta_o)\rho_N m_{21} - \frac{1}{2}(\rho-\rho_o)^2 m_{22} - \frac{1}{2}(\theta-\theta_o)^2\rho_N^2 m_{23}\ .\end{aligned}$$

Because $\partial P/\partial\rho=-W_*$ and $\partial P/\partial\theta=-\rho_* G_*$, when we derive $Q(\gamma)$ with respect to ρ and θ, we obtain:

$$\frac{\partial Q}{\partial\rho} = -\lim_{N\to\infty}\frac{1}{N}\mathrm{tr}[P^{-1}W_*] + m_{11} - (\theta-\theta_0)\rho_* m_{21} - (\rho-\rho_0)m_{22}$$

(71)

$$\frac{\partial Q}{\partial\theta} = -\rho_*\lim_{N\to\infty}\frac{1}{N}\mathrm{tr}[P^{-1}G_*] + \rho_* m_{12} - (\rho-\rho_0)\rho_* m_{21} - (\theta-\theta_0)\rho_*^2 m_{23}\ .$$

At $\rho=\rho_0$ and $\theta=\theta_0$, $P=P_0$ and so the first terms in these two equation converge to $-m_{11}$ and $-\rho_* m_{12}$ respectively, which proves that $\partial Q/\partial\gamma=0$ at $\gamma=\gamma_0$. The proof of consistency will be done after we prove that this point is in fact a local maximum. We thus need to show that the matrix of second derivatives $\partial^2 Q/\partial\gamma\partial\gamma'$ is negative definite in a neighborhood of γ_0.

Taking the derivative of (71) with respect to γ gives:

$$\frac{\partial^2 Q(\gamma)}{\partial\rho^2} = -\lim_{N\to\infty}\frac{1}{N}\mathrm{tr}[P^{-1}W_*P^{-1}W_*] - \lim_{N\to\infty}\frac{1}{N}\mathrm{tr}[P_0^{-1}W_*'W_*P_0^{-1}]$$

$$\frac{\partial^2 Q(\gamma)}{\partial\rho\partial\theta} = -\rho_*\lim_{N\to\infty}\frac{1}{N}\mathrm{tr}[P^{-1}W_*P^{-1}G_*] - \rho_*\lim_{N\to\infty}\frac{1}{N}\mathrm{tr}[P_0^{-1}W_*'G_*P_0^{-1}]$$

$$\frac{\partial^2 Q(\gamma)}{\partial\theta^2} = -\rho_*^2\lim_{N\to\infty}\frac{1}{N}\mathrm{tr}[P^{-1}G_*P^{-1}G_*] - \rho_*^2\lim_{N\to\infty}\frac{1}{N}\mathrm{tr}[P_0^{-1}G_*'G_*P_0^{-1}]\ .$$

Recalling that $P^{-1}P^{-1\prime}=V$ [which implies $PVP'=I$ and $P^{-1}=VP'$] and using a general notation for ρ and θ which are γ_i's $i=1,2$; the element i,j of this matrix of second derivatives, when evaluated at γ_0, may be written:

$$\frac{\partial^2 Q(\gamma)}{\partial\gamma_i\partial\gamma_j} = -\lim_{N\to\infty}\frac{1}{N}\{\mathrm{tr}[\partial P/\partial\gamma_j\ P^{-1}\ \partial P/\partial\gamma_i\ P^{-1}]+\mathrm{tr}[\partial P/\partial\gamma_j\ V\ \partial P'/\partial\gamma_i]\}$$

$$= -\lim_{N\to\infty}\frac{1}{N}\{\mathrm{tr}[\partial P/\partial\gamma_j\ VP'\partial P/\partial\gamma_i\ VP']+\mathrm{tr}[\partial P/\partial\gamma_j\ V\ \partial P'/\partial\gamma_i\ PVP']\}$$

$$= -\lim_{N\to\infty}\frac{1}{N}\{\mathrm{tr}[P'\partial P/\partial\gamma_j\ V\ (P'\partial P/\partial\gamma_i + \partial P'/\partial\gamma_i\ P\ )\ V]\}$$

$$= -\lim_{N\to\infty}\frac{1}{N}\mathrm{tr}[P'\partial P/\partial\gamma_j V\partial P'P/\partial\gamma_i V] = -\lim_{N\to\infty}\frac{1}{N}\mathrm{tr}[\partial P'/\partial\gamma_j PV\partial P'P/\partial\gamma_i V]$$

$$= -\frac{1}{2}\lim_{N\to\infty}\frac{1}{N}\mathrm{tr}[\ \partial P'P/\partial\gamma_j\ V\ \partial P'P/\partial\gamma_i\ V\ ] \qquad i,j = 1,2.$$

This result extends to the multi-parameter case straightforwardly. Using a well-known property of the vec operator, for a L-vector γ, the matrix of second derivatives $\partial Q(\gamma)/\partial\gamma\partial\gamma'$ evaluated at γ_o may be shown to be [see Magnus (1978)]:

$$(72)\quad \frac{\partial Q(\gamma)}{\partial\gamma\partial\gamma'} = -\frac{1}{2}\lim_{N\to\infty}\frac{1}{N}\frac{\partial \mathrm{vec}(P'P)'}{\partial\gamma}\left(V\otimes V\right)\frac{\partial \mathrm{vec}(P'P)}{\partial\gamma'}\,.$$

Since in the stable interval of ρ, V is positive definite everywhere, equation (72) implies a strict local maximum if the first derivatives are linearly independent. In other words, the parameters γ must be identifiable.

To give an example, the parameters in $w_{ij} = d_{ij}^{-(\alpha+\kappa)}$ would not be identifiable. We have proved that with an identifiable specification in the structure of contiguity, the ML estimators $\hat{\gamma}$ of the stable SAR(1) model are consistent. The consistency of $\hat{\beta}$ and $\hat{\sigma}^2$ follows in an obvious way.

5.3) Asymptotic Normality of SAR(1) ML Estimation

The standard way of proving asymptotic normality is to show by applying one version of existing limit central theorems CLT the asymptotic normality of $\sqrt{N}\,(\partial\ell_N/\partial\gamma)$ evaluated at γ_o. Our work on this subject, which is too partial to be included here seems to indicate that this approach could also be applied here. The main complication is caused by the fact that the observations are non identically and non independently distributed which makes difficult the application of a CLT theorem.

Using an approach that leads to conditions difficult to verify in practice, Magnus (1978) studied the asymptotic properties of ML estimations for GLS models with unknown parameters in the disturbance covariance matrix. Since the SAR(1) is a model that belongs to this class, for the conditions under which the estimators of the SAR(1) are asymptotically normally distributed, the reader is urged to consult this reference.

Section B) Maximum Likelihood Procedure

In the ensuing sections we will consider different aspects related to the estimation of the following generalized SAR(1) formulation:

(1) $y = X\beta + \epsilon$

(2) $\epsilon = \rho W\epsilon + u \qquad u \sim N(0,\sigma^2 I)$

with $w_{ij} = g(V_{ij}^1, V_{ij}^2, \ldots, V_{ij}^P; \theta_1, \theta_2, \ldots, \theta_H)$, i.e. a function of H parameters θ_h to be estimated and P pertinent variables V^p. In section 1) we describe a maximum likelihood procedure that aims to estimate the joint vector of unknown parameters $\eta = [\beta', \sigma^2, \rho, \theta']'$. In (2) W is assumed to be a symmetric matrix of contiguity.

Let $P = [I_N - \rho W]$ be the (N x N) matrix that depends on the parameter vector $\gamma = [\rho, \theta']'$. By equation (40) of section A, the log likelihood function is:

$$(3)\quad l(\beta', \sigma^2, \gamma') = -\frac{N}{2}\ln(2\pi) + \ln|P| - \frac{N}{2}\ln\sigma^2 - \frac{(y - X\beta)'V^{-1}(y - X\beta)}{2\sigma^2},$$

where $Var(\epsilon) = \sigma^2 V = \sigma^2 P^{-1}P^{-1}$, or also $V^{-1} = P'P$.

Remarks

1] Due to the fact that for each different value of the vector θ the matrix W changes, the simplification brought by replacing det P by $\prod(1 - \rho\lambda_i)$ is useless here. This means that at each iteration, the determinant of a (N x N) matrix has to be evaluated. By examining the following algorithm of estimation, one may observe that by exploiting on the computer the symmetric stockage mode of symmetrical weighting matrices, a substantial economy of memory space can be realized. This problem may become more and more important when the number of regions increases.

2] The results obtained in section A clearly suggested that if no normalization is applied on W, the possibility of finding a global maximum over the whole parametric space of the joint vector $[\beta', \sigma^2, \rho, \theta']'$ may become hazardous. In order to take advantage of the convenient properties induced by the normalization along with the space economy implied by the symmetrical stockage mode, the symmetrical normalization $\underset{\sim}{W}$ is preferred here.

The algorithm described in section 1) is formulated such that the user may choose to estimate the un-normalized version formed by (1) and (2) or the symmetrical normalized one given by equation (1) and:

$$(4)\quad \epsilon = \rho \underset{\sim}{W} \epsilon + u \qquad\qquad u \sim N(0, \sigma^2 I)$$

where $\underset{\sim}{w}_{ij} = \dfrac{w_{ij}}{\sqrt{\sum_i w_{ij}}\ \sqrt{\sum_j w_{ij}}}$.

1) M.L. estimation of the generalized SAR(1) residual

We first describe the ML procedure of the SAR(1) residual for the case with unormalized weighting matrix W. Following this, we provide the modifications needed when the weighting matrix is normalized symmetrical.

1.1) SAR(1) with un-normalized weighting matrix W

Maximizing the log likelihood function (3) with respect to β and σ^2 leads to the usual GLS expressions $\beta^* = (X'V^{-1}X)^{-1}X'V^{-1}y$ and also $\sigma^{*2} = (y - X\beta^*)'V^{-1}(y - X\beta^*)$. β^* and σ^{*2} are two explicit functions of the vector $\gamma = [\rho, \theta_1, \ldots, \theta_H]'$ that are such that the first order conditions corresponding to β and σ^2 are always satisfied. Replacing β and σ^2 in (3) by those values leads to the following concentrated log

likelihood function:

$$(5)\quad l^{**}(\gamma) = -\frac{N}{2}\ln(2\pi + 1) - \frac{N}{2}\ln\sigma^{*2} + \ln|P| \ .$$

Maximizing (5) with respect to γ_1 , $1 = 1, \ldots, H+1$ leads to:

$$(6)\quad \frac{\partial l^{**}}{\partial \gamma_1} = \mathrm{tr}\left(P^{-1}\frac{\partial P}{\partial \gamma_1}\right) - \frac{1}{\sigma^{*2}}(y - X\beta^*)'P'\frac{\partial P}{\partial \gamma_1}(y - X\beta^*) \ .$$

For (5) to be maximized, we must find the H+1 values $\hat{\gamma}_1$ such that the H+1 equations formed by (6) are all equal to zero simultaneously. At that value $\hat{\gamma}$ of the vector γ, the maximum likelihood estimators of β , σ^2 and γ are:

$\hat{\gamma}$, $\hat{\sigma}^2 = (y - X\hat{\beta})'\hat{V}^{-1}(y - X\hat{\beta})/N$ and $\hat{\beta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}y$.

Recalling that $P = [\,I - \rho W\,]$ and that γ_1 is either ρ or θ_h $h = 1,\ldots,H$, to evaluate equation (6) we must replace $\partial P/\partial\gamma_1$ by the appropriate expression in :

(7)

$\frac{\partial P}{\partial \gamma_1}$	ρ	θ_h
	$-W$	$-\rho\frac{\partial W}{\partial \theta_h}$

To calculate the variance covariance matrix of the joint vector $\hat{\eta}$ of ML estimators, we now calculate the elements of the matrix of second derivatives $\partial^2 l/\partial\eta\partial\eta'$. As is often done the variance covariance matrix of the estimates is calculated as:

$$-\left[\frac{\partial^2 l}{\partial\eta\partial\eta'}\right]^{-1}\Bigg|_{\eta = \hat{\eta}} \ .$$

Recalling that $\epsilon = y - X\beta$ and that $P\epsilon = u$, we then obtain:

$$(8)\quad \frac{\partial^2 l}{\partial\beta\partial\beta'} = -\frac{X'V^{-1}X}{\sigma^2}$$

(9) $\frac{\partial^2 l}{\partial\beta\partial\sigma^2} = -\frac{X'V^{-1}\epsilon}{\sigma^4}$

(10) $\frac{\partial^2 l}{\partial\beta\partial\gamma_1} = \frac{1}{\sigma^2}X'\left[P'\frac{\partial P}{\partial\gamma_1} + \frac{\partial P'}{\partial\gamma_1}P\right]\epsilon$

(11) $\frac{\partial^2 l}{\partial\sigma^4} = \frac{N}{2\sigma^4} - \frac{\epsilon'V^{-1}\epsilon}{\sigma^6}$

(12) $\frac{\partial^2 l}{\partial\sigma^2\partial\gamma_1} = \frac{1}{\sigma^4}u'\frac{\partial P}{\partial\gamma_1}\epsilon$

(13) $\frac{\partial^2 l}{\partial\gamma_1\partial\gamma_j} = tr\left[-P^{-1}\frac{\partial P}{\partial\gamma_j}P^{-1}\frac{\partial P}{\partial\gamma_1} + P^{-1}\frac{\partial^2 P}{\partial\gamma_1\partial\gamma_j}\right]$

$$-\frac{\epsilon'}{\sigma^2}\left[\frac{\partial P'}{\partial\gamma_j}\frac{\partial P}{\partial\gamma_1} + P'\frac{\partial^2 P}{\partial\gamma_1\partial\gamma_j}\right]\epsilon .$$

$\frac{\partial P}{\partial\gamma_j}$ may be found in (7). For $\frac{\partial^2 P}{\partial\gamma_1\partial\gamma_j}$, we have:

	γ_j	
$\frac{\partial^2 P}{\partial\gamma_1\partial\gamma_j}$	ρ	θ_g
ρ	0	$-\frac{\partial W}{\partial\theta_g}$
θ_h	$-\frac{\partial W}{\partial\theta_h}$	$-\rho\frac{\partial^2 W}{\partial\theta_h\partial\theta_g}$

1.2) SAR(1) with symmetrical normalized weighting matrix $\underset{\sim}{W}$

In this section we present the modifications that we must apply to the algorithm of section 1.1) in order to take account of a symmetrical normalized weighting matrix $\underset{\sim}{W}$. Using matrix algebra, this matrix can be written:

(14) $\underset{\sim}{W} = D\,W\,D'$,

where $D = D'$ is a diagonal matrix such that:

$$D = \begin{bmatrix} (\sum_{j=1}^{N} w_{1j})^{-\frac{1}{2}} & 0 & \cdots & 0 \\ 0 & (\sum_{j=1}^{N} w_{2j})^{-\frac{1}{2}} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & (\sum_{j=1}^{N} w_{Nj})^{-\frac{1}{2}} \end{bmatrix}$$

First derivatives

$$(15)\quad \frac{\partial \underset{\sim}{W}}{\partial \theta_h} = \frac{\partial D}{\partial \theta_h} W D' + D \frac{\partial W}{\partial \theta_h} D' + D W \frac{\partial D'}{\partial \theta_h} .$$

Let H be a diagonal matrix with elements h_{ii} equal to:

$$h_{ii} = \frac{\sum_{j=1}^{N} \frac{\partial w_{ij}}{\partial \theta_h}}{\sum_{j=1}^{N} w_{ij}}$$

It is then easy to prove that:

$$(16)\ \frac{\partial D}{\partial \theta_h} = \frac{\partial D'}{\partial \theta_h} = H D = D' H' \quad .$$

By replacing (16) in (15) one finally obtains:

$$(17)\ \frac{\partial \underset{\sim}{W}}{\partial \theta_h} = H \underset{\sim}{W} + D \frac{\partial \underset{\sim}{W}}{\partial \theta_h} D' + \underset{\sim}{W} H' \quad .$$

second derivatives

Call G the diagonal matrix define as H except that the derivative is taken with respect to θ_g instead. Of course we then have:

$$\frac{\partial D}{\partial \theta_g} = G D = D' G'$$

which allows us to obtain :

$$(18)\ \frac{\partial^2 \underset{\sim}{W}}{\partial \theta_h \partial \theta_g} = \frac{\partial H}{\partial \theta_g} \underset{\sim}{W} + H \frac{\partial \underset{\sim}{W}}{\partial \theta_g} + \underset{\sim}{W} \frac{\partial H'}{\partial \theta_g} + \frac{\partial \underset{\sim}{W}}{\partial \theta_g} H'$$

$$+ G D \frac{\partial \underset{\sim}{W}}{\partial \theta_h} D' + D \frac{\partial^2 \underset{\sim}{W}}{\partial \theta_h \partial \theta_g} D' + D \frac{\partial \underset{\sim}{W}}{\partial \theta_h} D' G' \quad .$$

More details on how this expression may be computed efficiently in the program of estimation may be found in Bolduc (1985).

section C) AN APPLICATION

In his 1968 paper "Accessibility and the Spatial Structure of the Irish Economy", P.M. O'Sullivan wants to test the existence of a functional relationship between the geographical configuration of the Irish economy in 1960 and the spatial structure of the existing transportation infrastructure. To do so, he regressed many indicators of the intensity of economic activity by county on different accessibility measures. To empirically characterize the transportation networks, O'Sullivan defined for each of the 26 counties a relative accessibility index :

$$A(i,R) = \sum_{j=1}^{N} d_{ij} \quad , i=1,\ldots,N,$$

where d_{ij} is a metric of distance between the centroids of county(area) i and j for the network R.

Among the different measures defined, O'Sullivan principally retained the arterial road accessibility index called ARA. His preference was based on the fact that at the beginning of the 60's most merchandise and passenger travel was done by road.

> "One problem which arises is which network to measure accessibility on. ... Attention was concentrated on the arterial road network proposed in 1964 as a result of a traffic census. This consists of the major links in the currently designated trunk road system. ... Railways carry a very small proportion of freight and passenger movements in Ireland and so position on the rail network may be deemed not so important as position on the road network. In 1960, 79 per cent of freight ton mileage in the republic was road-borne and in 1964 83 per cent. Thus road is becoming more dominant. In 1961, 93 per cent of passenger mileage was on road."

To represent the geographical variation of the economic activity, at the aggregated level, O'Sullivan used per capita personal income by county, which includes the income from all sectors of the economy excluding transfer of payment, elderly incomes and immigration. The regression results clearly indicate that the more accessible counties (low ARA index) [the south-east part of Ireland, see figure 5] have a high per capita income.

In the decomposition by economic sectors, O'Sullivan finally considers the Agricultural sector. In 1961, this sector was the main employer in Ireland. O'Sullivan noted the presence of a duality in the agricultural economy of Ireland. A subsistence agriculture characterized by small size farms can be found in the

less accessible areas [the north-west part] . The large scale production farms are mainly found in the south-east part of the country. To reflect this phenomenon, O'Sullivan used the on-farm consumption of produce as a percentage of gross agricultural output per county to measure the economic activity . The sign of the ARA in this regression is of course expected to be positive. Indeed the more a county is considered "far or isolated" the more important the on farm consumption should be. The results of this regression are given below:

$$(1)\quad y = \underset{(-2.75)}{-8.45} + \underset{(7.73)}{.005}\ ARA$$

$$R^2 = .697 \qquad \hat{\sigma}^2 = 12.54 \qquad \mathcal{L} = -32.387.$$

The figures in parentheses represent the t-statistics. A graphical representation of these two variables may be found in figure 1. Since y is a percentage [y constrained to the interval 0 to 1], a more appropriate procedure would have been to specify an aggregate logit form such as $\ln(y_t/1-y_t) = \alpha + \beta X_t + v_t$. In spite of that, to compare our results with those of many other spatial autocorrelation studies that used this regression case, we will study the problem of spatial autocorrelation of the linear model defined in equation 1. By plotting the residuals of this regression on figure 13, one can clearly detect a problem of positive spatial autocorrelation. Indeed, residuals of the same sign seem to be bunched together. This structure is so remarkable that recently, many spatial researchers have studied the econometric problems of this regression. Those efforts were mainly concentrated on the specification and the estimation under spatially autocorrelated residuals.

With respect to specification, one has to find the factors that can explain the spatial structure of the residual of figure 13. In that effect, Cliff and Ord (1981) have tried many different specifications as shown in section 8.6 of their monograph. Even though those specifications were potentially quite reasonable their empirical results were disappointing. Assuming that specification (1) as correct, they estimated the following SAR(1) formulation:

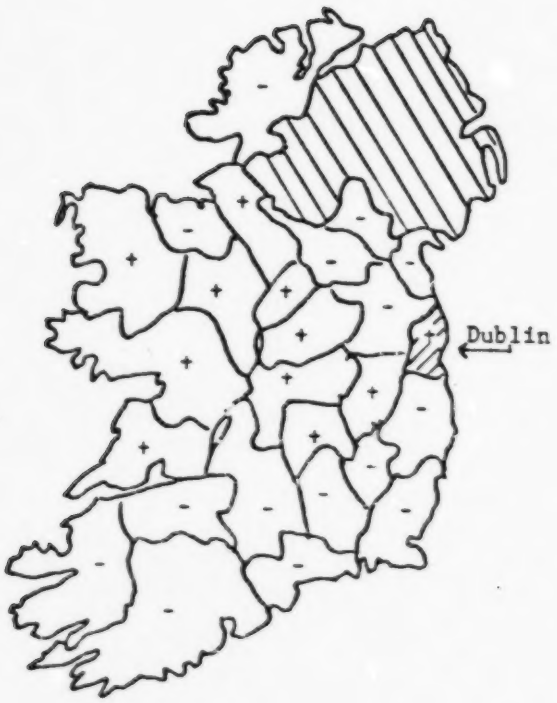
$$(2)\quad y = \alpha + \beta\ ARA + \epsilon$$

$$(3)\quad \epsilon = \rho M\epsilon + u,$$

where $u \sim N(0, \sigma^2 I_N)$, M is the normalized matrix defined such that $\sum_j^N m_{ij} = 1\ \forall i$, with $m_{ij} = w_{ij} / \sum_{j=1}^{N} w_{ij}$, where also:

$$w_{ij} = d_{ij}^{-\theta_1}\ \beta_{i(j)}^{\theta_2} \qquad \forall i \neq j\ ,$$

Figure 13



and where θ_1 and θ_2 are fixed to 1 a priori. d_{ij} is the distance measure and $\beta_{i(j)}$ represents the contact proportion between areas i and j. When the two areas are not neighbors of order one, this index is set to zero.

The mean deviation estimation results give:

$$(4)\quad y-\bar{y} = \underset{(4.286)}{.003}\ (ARA - \overline{ARA})$$

$$R^2 = .860 \quad \hat{\rho} = \underset{(5.2)}{.78} \qquad \hat{\sigma}^2 = 5.73 \quad \ell = NA$$

Even if the spatial autocorrelation coefficient is significant this procedure was unable to remove the positive spatial autocorrelation in the residuals. In a spatial common factor analysis, Burridge (1981), Blommestein (1983) and Bivand (1984) categorized equations (1) and (2) as a constrained model of the regressive/autoregressive spatial model:

$$(5)\quad y = \tilde{\alpha} + \beta\ ARA + \gamma M\ ARA + \rho M\ y + u\ ,$$

since if we impose the constraint $\gamma = -\rho\beta$ (common factor hypothesis), equation (5) can be written as:

$$(6)\quad (I - \rho M)\ y = \tilde{\alpha} + \beta\ (I - \rho M)\ ARA + u\ .$$

Positing $\tilde{\alpha}\iota = \alpha(I - \rho M)\iota = \alpha(1 - \rho)\iota$, equation (6) is merely another way of writing (2) and (3).

Their statistical results indicate that $\gamma \neq -\rho\beta$. They then conclude that the model estimated by Cliff and Ord is too constrained. However, even with this improved formulation, the unexplained factors are still correlated. Since we can assume that the SAR(1) formulation is almost as good as the regressive autoregressive formulation, we will now concentrate our effort on the estimation of a generalized version of the SAR(1) formulation. In section 1) we present estimation results of the SAR(1) model with the different boolean contiguity matrices used to plot figures 6, 10 and 11 of section A. In section 2) the estimations with generalized weighting schemes are displayed.

1) ACCESSIBILITY AND THE SPATIAL STRUCTURE OF THE IRISH ECONOMY: SOME RESULTS WITH DIFFERENT BOOLEAN CONTIGUITY MATRICES.

In table 1 below, one can find the results of estimation of the SAR(1) version of model (1) where the contiguity matrices are in turn:

$$(7)\ w_{ij} = \begin{cases} 1 \text{ if } i \text{ and } j \text{ are contiguous} \\ 0 \text{ otherwise,} \end{cases}$$

m_{ij} the non symmetric normalized form of (7) and w_{ij} the symmetrical normalized version.

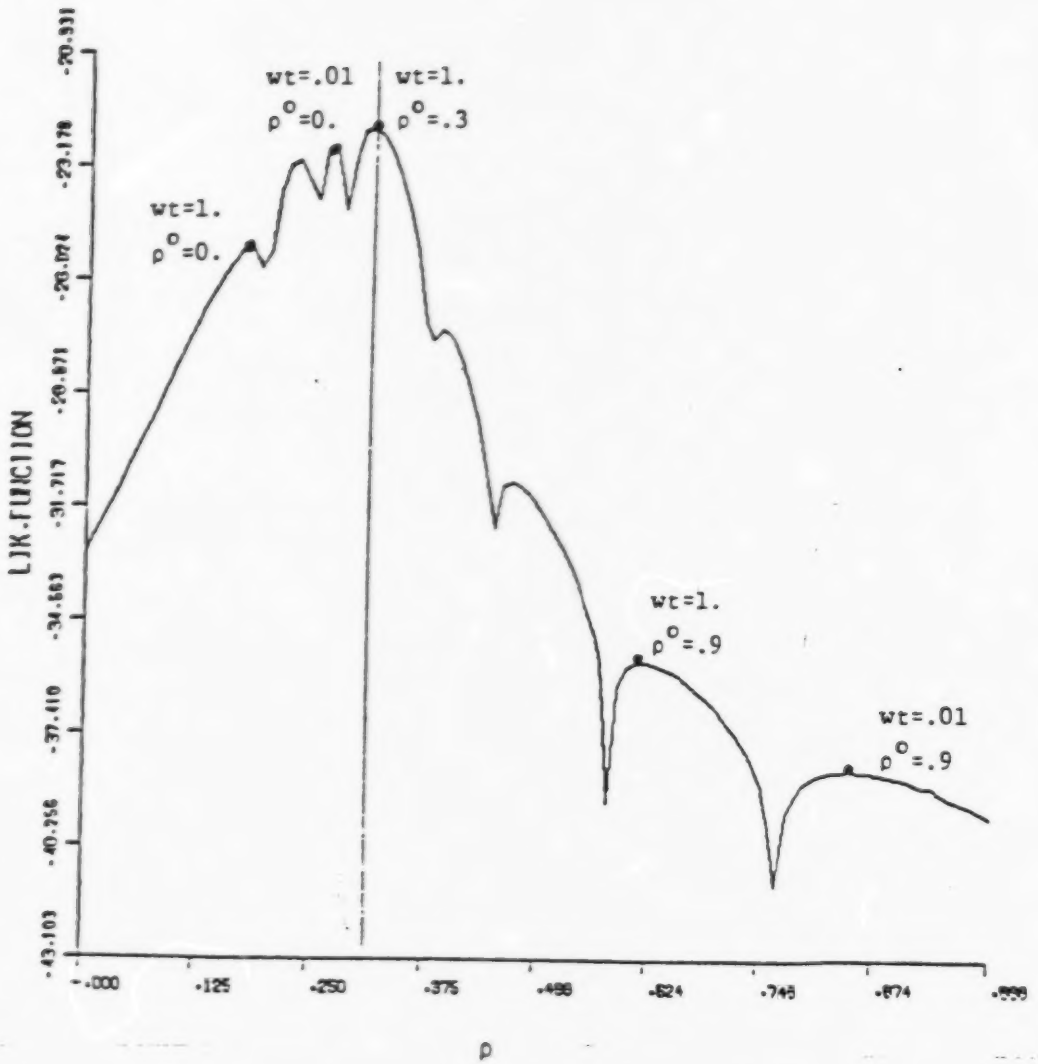
Table 1

Type of contiguity	α	β	ρ	R^2	σ^2	ℓ
[1] w_{ij} stable if $\lvert\rho\rvert<.195$	1.98 (0.89)	.003 (6.392)	.307 (17.55)	.95	4.17	-22.22
[2] m_{ij} stable if $\lvert\rho\rvert< 1.0$	4.67 (1.51)	.0024 (5.11)	.843 (9.03)	.86	5.86	-26.40
[3] w_{ij} stable if $\lvert\rho\rvert< 1.0$	5.29 (1.27)	.0026 (4.399)	.913 (13.63)	.82	4.65	-24.42

In figure 14 we restrict our study of the likelihood of the standard boolean case "figure 6" to the (0-1) domain. In accordance with the result established in section A, we find around the global maximum $\hat{\rho} = .307$ many other local optima. Some final estimations associated with the corresponding starting values ρ^0 and the weights required by the algorithm of estimation are indicated. It is noteworthy that the global maximum has been reached only once out of five attempts and that the starting value for that case was .3, a value very close to the final value. The results of this regression, as well as regressions [2] and [3] are as expected, with the coefficient of ARA being positive. The SAR(1) solution [1] is non stable since $\hat{\rho} > .195$.

As for estimation [1], the non symmetric m_{ij} version gives a non-significant constant term. The estimation of the slope is quite comparable to the one obtained in [1]. The significant value of $\hat{\rho}$ confirms the presence of positive spatial autocorrelation. The results presented in [2] are those of Bivand (1984).

Figure 14


wt=.01
ρ°=0.
wt=1.
ρ°=.3
wt=1.
ρ°=0.
wt=1.
ρ°=.9
wt=.01
ρ°=.9
LIK.FUNCTION
-.000
.125
.250
.375
.499
.624
.749
.874
ρ

The likelihood function was plotted in figure 10. The estimation results with a symmetrical normalized contiguity matrix is presented in [3]. By comparing with [2] we can see, that by imposing symmetry, the positive spatial correlation increased. Solution [3] was found in only 4 iterations. The GLS residuals were seen to show the presence of unexplained factors.

2) ESTIMATION OF THE PARAMETERS OF THE w_{ij}'s.

In this section we now try to estimate jointly with β, ρ and σ^2 the parameters of the non-negative functional forms:

$$w_{ij} = g(V_{ij}^1, V_{ij}^2, \ldots, V_{ij}^P, \theta_1, \theta_2, \ldots, \theta_H),$$

i.e. a function of H parameters and P pertinent variables. In the following, all estimations are performed with the symmetrical normalized form $\underset{\sim}{w}_{ij}$ of (7).

2.1) $w_{ij} = d_{ij}^{-\theta_1}$

Since the strength of the relation between two areas generally increases with their proximity, it is appropriate to give more weight to the closest areas. Since we do not need to restrict the dependency only to the contiguous areas, the functional form estimated here is:

(8) $w_{ij} = d_{ij}^{-\theta_1} \quad \forall\, i \neq j$ with $w_{ii} = 0\ \forall i$.

After being concentrated with respect to α, β and σ^2, the likelihood function is maximized only with respect to ρ and θ_1. To show how the estimation may become cumbersome if no normalization is made on the contiguity matrix, we present in figure 15 a grid over the region $.4<\rho<1$ and $.2<\theta_1<.5$. The accompanying table of values of the likelihood function shows many drastic changes. For example, for ρ fixed at .5498 [line 7], and θ_1 varying between .2 and .5, the likelihood function fluctuates between -21.8290 and -29.184. When the symmetrical normalization is used, the behavior of the likelihood function becomes much more uniform. The grid around the maximum likelihood solution $\hat{\rho} = .8005$ and $\theta_1 = 3.356$ is presented in figure 16. The smoothness of the corresponding likelihood function explains why the estimation algorithm took only nine iterations to find that solution. The ML results are:

(9) $y = \underset{(-.686)}{-3.035} + \underset{(5.698)}{0.004}\ ARA$

$R^2 = .801 \quad \underset{(5.2151)}{\hat{\rho} = .8005} \quad \hat{\sigma}^2 = 6.712 \quad \mathcal{L} = -27.3607$

with $\theta_1 = \underset{(3.738)}{3.356}$.

Figure 15

FMAXI,FMINI : -21.43162155007 -29.18395318164

1	.9990	-23.3127	-23.3589	-21.4316	-24.2673	-22.3621	-22.8969	-23.6631	-24.3935	-23.1083
2	.9241	-23.4767	-25.3998	-21.5557	-22.3569	-22.4342	-22.6883	-23.4193	-24.1642	-24.8760
3	.8493	-23.7134	-23.5922	-21.9230	-21.7351	-22.9126	-22.5101	-23.1437	-23.8924	-24.6125
4	.7744	-24.0264	-23.3540	-22.7087	-21.5246	-23.8497	-22.4508	-22.8303	-23.5696	-24.3020
5	.6995	-24.4097	-23.4636	-25.9476	-21.7039	-22.0414	-22.8212	-22.5948	-23.1972	-23.9263
6	.6246	-24.8671	-23.7872	-23.5770	-22.4562	-21.6083	-23.4387	-22.3635	-22.8143	-23.4744
7	.5498	-25.4015	-24.2828	-23.5229	-29.1840	-21.9163	-21.8290	-24.2630	-22.5991	-22.9784
8	.4749	-26.0183	-24.9326	-23.9759	-23.5568	-23.6186	-21.8182	-22.0022	-23.9082	-22.6721
9	.4000	-26.7254	-25.7292	-24.7482	-23.9228	-23.6374	-23.6626	-21.9325	-21.9273	-24.6375
		.2000	.2375	.2750	.3125	.3500	.3875	.4250	.4625	.5000

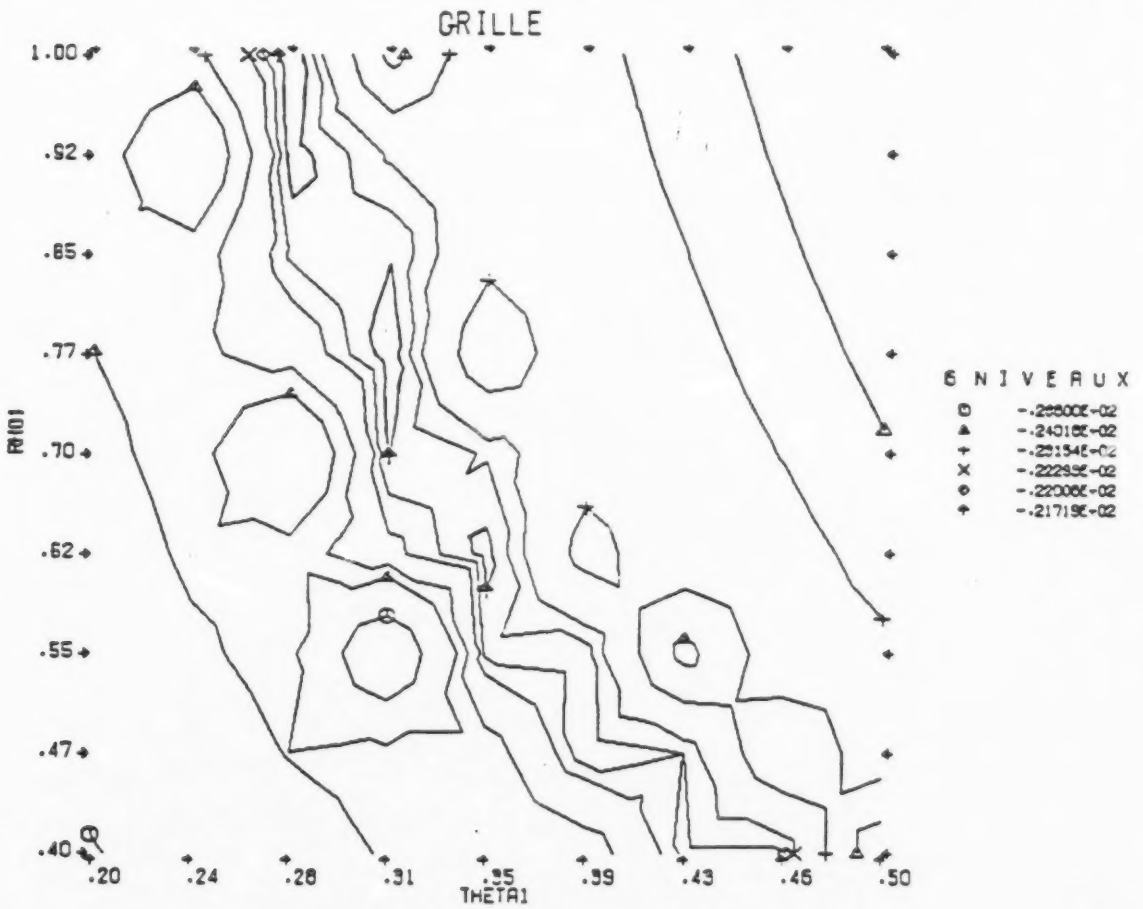
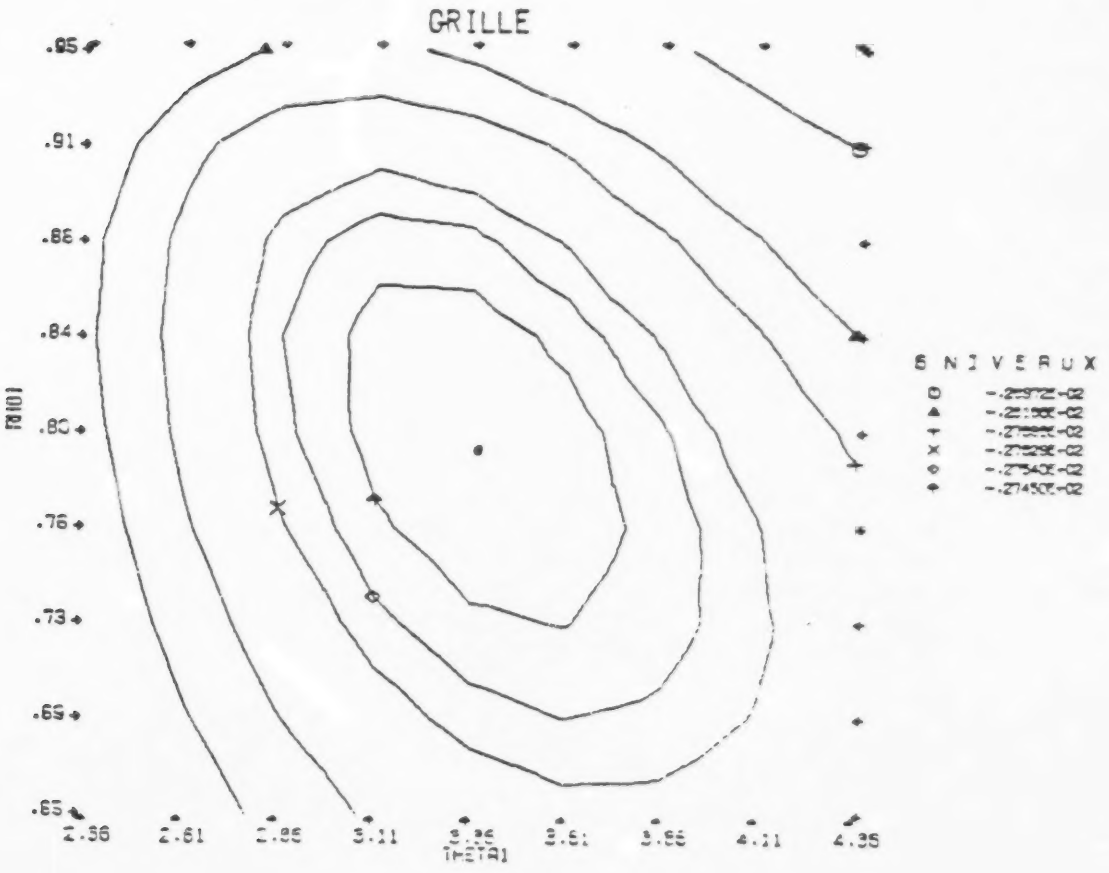


Figure 16

FMAXI,FMINI : -27.360482604873 -29.777808823134

1	.9505	-28.6792	-28.3144	-28.1341	-28.1039	-28.2360	-28.4954	-28.8552	-29.2899	-29.7778
2	.9130	-28.3415	-27.9635	-27.7434	-27.6773	-27.7496	-27.9389	-28.2216	-28.5732	-28.9803
3	.8755	-28.2182	-27.8304	-27.5872	-27.4847	-27.5089	-27.6411	-27.8405	-28.1473	-28.4845
4	.8380	-28.1990	-27.8040	-27.5422	-27.4095	-27.3938	-27.4783	-27.6448	-27.8737	-28.1552
5	.8005	-28.2426	-27.8442	-27.5674	-27.4101	-27.3607	-27.4048	-27.5261	-27.7090	-27.9393
6	.7630	-28.3222	-27.9215	-27.6440	-27.4662	-27.3887	-27.3984	-27.4811	-27.6223	-27.8078
7	.7255	-28.4332	-28.0346	-27.7399	-27.5663	-27.4636	-27.4463	-27.4956	-27.6010	-27.7509
8	.6880	-28.5993	-28.2058	-27.9076	-27.7020	-27.5829	-27.5393	-27.5603	-27.6246	-27.7516
9	.6505	-28.7632	-28.3798	-28.0813	-27.8682	-27.7344	-27.6709	-27.6680	-27.7154	-27.8538
		2.2543	2.6043	2.8543	3.1043	3.3543	3.6043	3.8543	4.1043	4.3543



The value of $\hat{\rho}$ confirms the presence of a strong positive spatial autocorrelation. The t value of θ_1 clearly indicates that the degree of relation between the residuals of the regression is a non linear inverse function of the distance. The sign of θ_1 is as expected. However, the spatial repartition of the GLS residuals, not presented here, shows no great improvement over the linear regression results of figure 13. Similar results(not reported here) have been obtained with formulations such as:

$$w_{ij} = (\ln d_{ij})^{-\theta_1} \quad \text{and} \quad w_{ij} = e^{-\theta_1 d_{ij}} .$$

According to the level of the likelihood function obtained in equation (9) and table 1, a formulation that takes account of contiguous regions only, seems more reasonable for this model. In order to test this hypothesis, a model with w_{ij} elements specified as:

$$(10)\quad w_{ij} = d_{ij}^{-\theta_1} (\theta_3 + \theta_2 g_{ij}) \qquad \theta_3 \geq 0 \text{ and } \theta_2 \geq 0$$

with $g_{ij} = 1$ if i and j are contiguous and $=0$ otherwise, was estimated. Because of the symmetrical normalization, one of the θ_2 or θ_3 is unidentified. (10) can then equivalently be written as:

$$(11)\quad w_{ij} = d_{ij}^{-\theta_1} (\theta_3 + g_{ij}) \text{ or } w_{ij} = d_{ij}^{-\theta_1} (1 + \theta_2 g_{ij}).$$

When $\theta_2 = 0$, the model of equation (9) is obtained. For $\theta_3 = 0$, the model with residual dependency limited to contiguous regions is obtained. The ML estimation of the unrestricted model (11) leads to a value of $\theta_3 = 0$. The likelihood ratio test computed in Bolduc (1985) classifies the model with dependency restricted to contiguous regions as better.

estimation results with dependency limited to contiguous regions

(12) $\gamma = 5.041 + 0.003\ ARA$
(1.177) (4.407)

$R^2 = .818 \quad \hat{\rho} = .909 \quad \hat{\sigma}^2 = 4.65 \quad \ell = -24.382$
(13.29)

with $\theta_1 = 0.282$.
(0.295)

In a formulation where the residuals depend only on the contiguous residuals, the distance d_{ij} with a t value of .295, loses its signification. When θ_1 is fixed to zero, the estimation [3] of table 1 is produced. This result clearly makes sense since the distance between zones and the contiguity variable g_{ij} in (10) are two indices of proximity. The estimations obtained in Bolduc (1985) and to some extent here indicate that when the contiguity is imposed, the distance has nothing left to explain. When the non-contiguous version is imposed, g_{ij} become non significant. It would be interesting to verify if this phenomenon is specific only to this study.

If after having introduced the distance in w_{ij}, their remains some spatial autocorrelation, other factors that explain the correlation must be added. Some suggestions are provided below.

Some examples of variables that may explain the spatial autocorrelation

Variables that can be included in w_{ij} are either double indices variables or variables created from a combination of socio-economic variables specific to zone i and zone j respectively. Examples of double indices variables are:

-number of trips between zone i and j, by car, bus or other mean of transportation
-the degree of utilization of telephonic lines
-an inverse function of transportation cost that usually take account of the travel time, waiting time etc...

Examples of the second kind of variables , includes:

-product or sum of the populations or incomes in the two areas concerned. [The more important the economic activity is, the more the link is expected to be strong.]

In our empirical case, many general forms that include the distance along with other factors were tried. From those attempts, only the model presented in 2.2) was found to be significant.

2.2) $w_{ij}=d_{ij}(\ \ln|rev_i - rev_j|\ )^{-\theta_1}\ (\ln|rev_i - rev_j|\ )^{-\theta_2}$ if $i\neq j$, $w_{ii}=0\ \forall i$

A justification for this form may be found in Bolduc (1985).

The estimation results are:

(13) $y = -.242 + 0.004\ ARA$
$\quad\quad (-.063)\quad (6.058)$

$R^2 = .87 \quad \hat{\rho} = .868 \quad \hat{\sigma}^2 = 5.77 \quad \ell = -26.591$
$\quad\quad\quad\quad\quad (8.89)$

with $\theta_1 = 4.005 \quad\quad \theta_2 = 0.532.$
$\quad\quad\quad (3.497) \quad\quad\quad (2.244)$

In this regression, $\hat{\rho}$, θ_1 and θ_2 are all significant and have the expected signs. Compared to the results obtained in (8), the likelihood function did not increase enough to justify the inclusion of that supplementary variable. The resulting spatial residual configuration was found to be even worse than in the linear regression case.

The other variables that were tried were all non significant. We must then conclude that even by estimating the parameters of the functional form w_{ij} , we were not able to break the spatial correlation in the residuals. Even if the results of estimation are not so encouraging, the generalized SAR(1) formulation was found to be easy to apply and implement. From our work, one can conclude that in any study that uses spatial data, if the researcher uses a normalized form such as the two forms studied in this paper, the vector of parameter $[\beta', \sigma^2, \rho, \theta']'$ will always be quite easily estimable if ρ is restricted to the stable interval $-1<\rho<1$.

CONCLUSION

Because we felt there was a need for it, we included in section A) of the paper some previous and new results related to the estimation of regressions with spatially autocorrelated residuals. The main purpose of this paper was to develop a method of estimation that deals with this problem. We have shown the feasibility of a generalize SAR(1) model that uses a special kind of normalized contiguity matrix. Proofs of consistency of estimation were given.

Extensions of the method developed in this paper which are presently considered by the author includes the following: - the estimation of travel demand with spatially autocorrelated residuals. (This type of data [origin / destination] requires a special treatement). A direct extension of the previous work and an error component form are two alternative formulations investigated. - the Discrete Choice Modeling with Spatially Autocorrelated Alternatives is also being investigated. This approach can potentially have an important impact on the modelization of choice of location .

As other direct extensions of the present work, one can consider the estimation of the weighting matrices in the following low order models: the spatial SMA(1) model of Haining (1978c), higher order SAR and SMA such as the SAR(2) of Brandsma and Ketellaper (1979a) and the not yet considered SARMA(1, 1).

NOTES

1. The stationarity conditions for cases 1 2 and 3 can be obtained using the theory of random field processes. Details may be found in Haining (1977) or (1978b). See also chapter 1 of Bolduc (1985).

2. See Abramowitz and Stegun (1972) section 22.16.

3. Vec(ABC) = $(A \otimes C')$ vec(B). This function stacks the rows of a (P x Q) matrix into a (PQ x 1) vector.

4. A detailed derivation for cases 4 and 5 may be found in Bolduc (1985).

5. for $I + \rho W + \rho^2 W^2 + \ldots$ to converge, $|\rho \lambda_{MAX}(W)| < 1$ which implies $|\rho| < 1/|\lambda_{MAX}(W)|$.

6. As observed by Fisher (1971), the concept of asymptoticity which is justifiable with serial data becomes questionable with spatial data. As he suggested, in contrast to time series data, it is difficult to admit that as $N \to \infty$ (number of areas increases), the region will grow in such a way that the addition of areas on the border will be consistent with the structure of dependency already present. Notwithstanding, the state of the art being what it is, we will use the large sample concept and its associated theorical simplifications.

7. In the following, the determinant of a matrix A is denoted det A. $|A|$ will stand for the absolute value of the determinant of A. The absolute value of a scalar y is also denoted $|y|$.

8. For that case, taking a lag one spatial difference is recommended see Martin (1974).

BIBLIOGRAPHY

Abramowitz, M. and I.A. Stegun, 1972, Handbook of Mathematical Functions, Dover, New York.

Amemiya, T; 1985, Advanced Econometrics, Harvard.

Anselin, L., 1982, "A Note on Small Sample Properties of Estimators in a First Order Spatial Autoregressive Model", "Environment and Planning", A, 14, 8, (August), 1023-30.

Bivand, R.S., 1984, "Regression Modeling with Spatial Dependance: an Application of Some Class Selection and Estimation Methods", Geographical Analysis, 16, 1 (January), 25-37.

Blommestein, H.J., 1983, "Specification and Estimation of Spatial Econometric Models. A Discussion of Alternative Strategies for Spatial Economic Modeling", Regional Science and Urban Economics, 13, 251-270.

Bodson, P. and D. Peeters, 1975, "Estimation of the Coefficients of a Linear Regression in the Presence of Spatial Autocorrelation. An Application to the Belgian Labour-Demand Function", Environment and Planning, A, 7, 455-472.

Bolduc, Denis, 1985, "Deux Procedures d'estimation en presence d'autocorrelation spatiale." These de doctorat, publiee par le Centre de Recherches sur les Transports, Universite de Montreal. (December)

Brandsma, A.S. and R.H. Ketellaper, 1979a, "A Biparametric Approach to Spatial Autocorrelation", Environment and Planning, A, 11, 51-58.

Brandsma, A.S. and R.H. Ketellaper, 1979b, "Further Evidence on Alternative Procedures for Testing of Spatial Autocorrelation Among Regression Disturbances", in Exploratory and Explanatory Statistical Analysis of Spatial Data, edited by Bartels, C.P.A. and R.H. Ketellaper, 1979, Martinus Nijhoff Publishing.

Burguete, J.F., A.R. Gallant and G. Souza, 1982, "On Unification of the Asymptotic Theory of Nonlinear Econometric Models", Econometric Reviews, 1, 2, 151-190.

Burridge, P., 1980, "On the Cliff-Ord Test for Spatial Correlation", Journal of the Royal Statistical Society, Series B, 42, 1, 107-8.

Burridge, P., 1981, "Testing for a Common Factor in a Spatial Autoregressive Model", Environment and Planning, A, 13, 795-800.

Cliff, A.D. and J.K. Ord, 1969, "The Problem of Spatial Autocorrelation" in London Papers in Regional Science, vol. 1, Studies in Regional Science, London, edited by Scott, A.J.

Cliff, A.D. and J.K. Ord, 1971, "Evaluating the Percentage Points of a Spatial Autocorrelation Coefficient", Geographical Analysis, 3, 51-62.

Cliff A.D. and J.K. Ord, 1973, Spatial Autocorrelation, Pion, London.

Cliff A.D. and J.K. Ord, 1975, "Space-Time Modelling with an Application to Regional Forecasting", Transaction of the Institute of British Geographers, 64, (March), 119-128.

Cliff A.D. and J.K. Ord, 1981, Spatial Processes, Models and Application, Pion, London.

Cochran, D. and G.H. Orcutt, 1949, "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms", Journal of the American Statistical Association, 44: 32-61.

Davidon, W.C., 1959, "Variable Metric Method for Minimization", A.E.C. Research and Development Re[port, ANL-5990.

Doreian, P., 1978, "On the Linear Models with Spatially Distributed Data", Chapter 11 in Sociological Methodology, edited by Leinhart, S.S., The American Sociological Association, School of Urban and Public Affairs, Carnegie-Mellon University, Pittsburgh.

Durbin, J., 1960, "Estimation of Parameters in Time-Series Regression Models", Journal of the Royal Statistical Society, Series B, 22, 1, 139-153.

Durbin, J. and G. S. Watson, 1950, "Testing for Serial Correlation in Least Squares Regression I", Biometrika, 37, 409-28.

Fisher, W. D., 1971, "Econometric Estimation with Spatial Dependence", Regional and Urban Economics, Operational Methods, 1, 1, 19-40.

Fletcher, R. and M. J. D. Powell, 1963, "A Rapidly Convergent Descent Method for Minimization", Computer5 Journal, 6, 163-68.

Fuller, W. A., 1976, Introduction to Statistical Time Series, John Wiley and Sons.

Gallant, R. A. and A. Holly, 1980, "Statistical Inference in an Implicit, Nonlinear, Simultaneous Equation Model in the Context of Maximum Likelihood Estimation", Econometrica, 48, 3, (April), 697-720.

Gatrell, A. C., 1977, "An Introduction to Spatial Autocorrelation and its Geographical Applications", Discussion Papers in Geography, number 2, (April), University of Salford, England.

Gatrell, A. C., 1979a, "Autocorrelation in Spaces", Environment and Planning, 11, 507-16.

Gatrell, A. C., 1979b, "The Autocorrelation Structure of Central-Place Populations in Southern Germany, in Statistical Application in the Spatial Sciences, Chapter 4, edited by Wrigley, N.

Geary, R. C., 1968, "The Contiguity Ratio and Statistical Mapping", in Spatial Analyses: A Reader in Statistical Geography, edited by Berry, B. L. J. and D. F. Marble, Englewood Cliffs, N. J., Prentice-Hall.

Gourieroux, C., A. Monfort and A. Trognon, 1984, "Pseudo Maximum Likelihood Methods: Theory", Econometrica, 52, 3, (May), 681-701.

Granger, C. W. J., 1969, "Spatial Data and Time Series Analysis", in London Papers in Regional Science, vol. 1, Studies in Regional Science, London, edited by Scott, A. J.

Griffith, D., 1980, "Towards a Theory of Spatial Statistics", Geographical Analysis, 12, 4, 325-39.

Haining, R. P., 1977, "Model Specification In Stationary Random Fields", Geographical Analysis, IX, 107-29.

Haining, R. P., 1978a, "Estimating Spatial Interaction Models", Environment and Planning, A, 10, 305-20.

Haining, R. P., 1978b, "Specification and Estimation Problems in Model of Spatial Dependence", Northwestern University Studies in Geography, Evanston.

Haining, R. P., 1978c, "The Moving Average Model for Spatial Interaction", Transaction of the Institute of British Geographers, 68, 4, (December), 493-504.

Haining, R. P., 1978d, "A Spatial Model for High Plains Agriculture", Annals of the Association of American Geographers, 68, 4 (December), 493-504.

Haining, R. P., 1980, "Spatial Autocorrelation Problems", in Geography and the Urban Environment, Progress in Research and Application, edited by Herbert, D. T. and R. J. Johnston, John Wiley and Sons, chapter 1.

Hepple, L. W., 1976, "A Maximum Likelihood Model for Econometric Estimation with Spatial Series", in Theory and Practice in Regional Science, edited by Masser, I., Pion, 90-104.

Hepple, L. W., 1979, "Bayesian Analysis of the Linear Model with Spatial Dependance", in Exploratory and Explanatory Statistical Analysis of Spatial Data, edited by Bartels, C. P. A. and R. H. Ketellaper, Martinus Nijhoff Publishing.

Hordijk, L., 1974, "Spatial Correlation in the Disturbances of a Linear Interregional Model", Regional and Urban Economics, 4, 117-140.

Hordijk, L., 1979, "Use of Econometric Relations in Space", Papers of the Regional Science Association, 42, 100-115.

Hordijk, L. and J. H. P. Paelinck, 1976, "Some Principles and Results in Spatial Econometrics", Recherches economiques de Louvain, 42, 3, 175-97.

Jenrich, R. I., 1969, "Asymptotic Properties of Non-Linear Least Squares Estimators", The Annals of Mathematical Statistics, 40, 2, 633-43.

Johnston, J., 1984, Econometric Methods, 3rd Edition, McGraw-Hill Book Co.

Judge, G. G., W. E. Griffiths, R. C. Hill, and T. C. Lee, 1980, The Theory and Practice of Econometrics, John Wiley and Sons.

Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lutkepohl and T.C. Lee, 1982, Introduction to the Theory and Practice of Econometrics, John Wiley and Sons.

Magnus, J.R., 1978, "Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix", Journal of Econometrics, 7, 281-312.

Martin, R.L., 1974, "On Spatial Dependence, Bias and the Use of First Spatial Differences in Regression Analysis", AREA, 6, 3, 185-194.

Monfort, A., 1982, Cours de Statistique Mathematique, Economica.

Oden, N.L., 1984, "Assessing the Significance of a Spatial Correlogram", Geographical Analysis, 16, 1, 1-16.

Ord, J.K., 1975, "Estimation Methods for Models of Spatial Interaction", Journal of the American Statistical Association, 70, 349, 120-26.

Ord, J.K., 1981, "Towards a Theory of Spatial Statistics: A Comment", Geographical Analysis", 13, 1, 86-91.

O'Sullivan, P.M., 1968, "Accessibility and the Spatial Structure of the Irish Economy", Regional Studies, 2, 195-206.

Theil, H., 1971, Principles of Econometrics, John Wiley and Sons.

White, H., 1982, "Maximum Likelihood Estimation of Misspecified Models", Econometrica, 50, 1, (January), 1-25.

Whittle, P., 1954, "On Stationary Processes in the Plane", Biometrika, 41, 434-49.